

# Implication of genetic variants in primary microRNA processing sites in the risk of multiple sclerosis



Michael Hecker,<sup>a\*</sup> Brit Fitzner,<sup>a</sup> Elena Putscher,<sup>a</sup> Margit Schwartz,<sup>a</sup> Alexander Winkelmann,<sup>a</sup> Stefanie Meister,<sup>a</sup> Ales Dudesek,<sup>a</sup> Dirk Koczan,<sup>b</sup> Peter Lorenz,<sup>b</sup> Nina Boxberger,<sup>a,1</sup> and Uwe Klaus Zettl<sup>a,1</sup>

<sup>a</sup>Department of Neurology, Division of Neuroimmunology, Rostock University Medical Center, Gehlsheimer Str. 20, Rostock 18147, Germany

<sup>b</sup>Institute of Immunology, Rostock University Medical Center, Schillingallee 70, Rostock 18057, Germany

## Summary

**Background** Multiple sclerosis (MS) is a chronic inflammatory disease of the central nervous system with a well-established genetic contribution to susceptibility. Over 200 genetic regions have been linked to the inherited risk of developing MS, but the disease-causing variants and their functional effects at the molecular level are still largely unresolved. We hypothesised that MS-associated single-nucleotide polymorphisms (SNPs) affect the recognition and enzymatic cleavage of primary microRNAs (pri-miRNAs).

**Methods** Our study focused on 11 pri-miRNAs (9 primate-specific) that are encoded in genetic risk loci for MS. The levels of mature miRNAs and potential isoforms (isomiRs) produced from those pri-miRNAs were measured in B cells obtained from the peripheral blood of 63 MS patients and 28 healthy controls. We tested for associations between SNP genotypes and miRNA expression in *cis* using quantitative trait locus (*cis*-miR-eQTL) analyses. Genetic effects on miRNA stem-loop processing efficiency were verified using luciferase reporter assays. Potential direct miRNA target genes were identified by transcriptome profiling and computational binding site assessment.

**Findings** Mature miRNAs and isomiRs from hsa-mir-26a-2, hsa-mir-199a-1, hsa-mir-4304, hsa-mir-4423, hsa-mir-4464 and hsa-mir-4492 could be detected in all B-cell samples. When MS patient subgroups were compared with healthy controls, a significant differential expression was observed for miRNAs from the 5' and 3' strands of hsa-mir-26a-2 and hsa-mir-199a-1. The *cis*-miR-eQTL analyses and reporter assays pointed to a slightly more efficient Drosha-mediated processing of hsa-mir-199a-1 when the MS risk allele T of SNP rs1005039 is present. On the other hand, the MS risk allele A of SNP rs817478, which substitutes the first C in a CNNC sequence motif, was found to cause a markedly lower efficiency in the processing of hsa-mir-4423. Overexpression of hsa-mir-199a-1 inhibited the expression of 60 protein-coding genes, including *IRAK2*, *MIF*, *TNFRSF12A* and *TRAF1*. The only target gene identified for hsa-mir-4423 was *TMEM47*.

**Interpretation** We found that MS-associated SNPs in sequence determinants of pri-miRNA processing can affect the expression of mature miRNAs. Our findings complement the existing literature on the dysregulation of miRNAs in MS. Further studies on the maturation and function of miRNAs in different cell types and tissues may help to gain a more detailed functional understanding of the genetic basis of MS.

eBioMedicine 2022;80:  
104052

Published online xxx

[https://doi.org/10.1016/j.](https://doi.org/10.1016/j.ebiom.2022.104052)

[ebiom.2022.104052](https://doi.org/10.1016/j.ebiom.2022.104052)

**Abbreviations:** ANOVA, analysis of variance; APC, allophycocyanin; bp, base pairs; BV, brilliant violet; CNS, central nervous system; C<sub>T</sub>, threshold cycle; EAE, experimental autoimmune encephalomyelitis; EBV, Epstein-Barr virus; EDSS, Expanded Disability Status Scale; EDTA, ethylenediaminetetraacetic acid; eQTL, expression quantitative trait locus; EUR, European; FITC, fluorescein isothiocyanate; GEO, Gene Expression Omnibus; GLuc, *Gaussia* luciferase; GWAS, genome-wide association study; h, hours; HGNC, HUGO Gene Nomenclature Committee; HLA, human leukocyte antigen; IRT, immune reconstitution therapy; isomiR, microRNA isoform; kb, kilobase; LD, linkage disequilibrium; MAF, minor allele frequency; mfe, minimum free energy; miR/MIR/miRNA, microRNA; mRNA, messenger RNA; MS, multiple sclerosis; mv, missing values; n/N, number; ng, nanogram; nt, nucleotide; OR, odds ratio; PBMC, peripheral blood mononuclear cells; PE, phycoerythrin; PerCP, peridinin-chlorophyll; PPMS, primary progressive multiple sclerosis; pri-miRNA, primary microRNA; qPCR, quantitative polymerase chain reaction; RA, risk allele; RPM, reads per million; RRMS, relapsing-remitting multiple sclerosis; RT, reverse transcription; SD, standard deviation; SEAP, secreted alkaline phosphatase; SNP, single-nucleotide polymorphism; SPMS, secondary progressive multiple sclerosis; sQTL, splicing quantitative trait locus; TAC, Transcriptome Analysis Console; TC, transcript cluster; T<sub>H</sub>17, T helper 17 cell; UNG, uracil-N-glycosylase; UTR, untranslated region

\*Corresponding author.

E-mail address: [michael.hecker@rocketmail.com](mailto:michael.hecker@rocketmail.com) (M. Hecker).

<sup>1</sup> These authors contributed equally to this work.

**Funding** This study was funded by the Rostock University Medical Center (FORUN program, grant: 889002), Sanofi Genzyme (grant: GZ-2016-11560) and Merck Serono GmbH (Darmstadt, Germany, an affiliate of Merck KGaA, CrossRef Funder ID: 10.13039/100009945, grant: 4501860307). NB was supported by the Stiftung der Deutschen Wirtschaft (sdw) and the FAZIT foundation. EP was supported by the Landesgraduiertenförderung Mecklenburg-Vorpommern.

**Copyright** © 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

**Keywords:** Multiple sclerosis; Genetic risk; Single-nucleotide polymorphisms; MicroRNAs; B cells; Expression quantitative trait loci; Primary microRNA processing; MicroRNA target genes; Immune reconstitution therapy

### Research in context

#### *Evidence before this study*

Genome-wide association studies of multiple sclerosis (MS) have identified over 200 risk loci across the human genome. However, our understanding of disease architecture is still far from complete. Discerning the causal variants and gaining functional insights are essential steps towards bridging the gap between genotypes and clinical phenotypes. Some disease susceptibility loci encode microRNAs (miRNAs). We suspected that the genotype of MS-associated single-nucleotide polymorphisms (SNPs) might influence the enzymatic processing and thus the expression of such miRNAs and their target genes.

#### *Added value of this study*

We used B cells from the peripheral blood of MS patients and healthy controls to measure selected mature miRNAs and isoforms produced from primary miRNA transcripts that originate from genetic risk loci for MS. We found that miRNAs from both the 5' and the 3' strand of hsa-mir-26a-2 and hsa-mir-199a-1 were differentially expressed between the groups. A higher expression of miR-199a miRNAs was observed in the presence of the MS risk allele of SNP rs1005039, whereas the risk allele of SNP rs817478 was associated with a lower expression of hsa-miR-4423-5p. These genetic effects on primary miRNA processing were verified using reporter assays. Moreover, we identified known and new potential direct target genes of these miRNAs.

#### *Implications of all evidence available*

We show that MS-associated SNPs in primary miRNA processing sites can affect the expression of mature miRNAs. These findings contribute to a better understanding of the molecular mechanisms underlying the pathogenesis of MS. Further cell type-specific studies on transcriptional regulation and RNA processing are necessary to elucidate the functional nature of the numerous genetic determinants of disease susceptibility. These efforts will help to define the specific cell types and complex regulatory networks causally involved in MS.

### Introduction

Multiple sclerosis (MS) is a chronic, inflammatory and neurodegenerative disease of the central nervous system (CNS).<sup>1</sup> The pathological hallmark of MS is the formation of demyelinating lesions in the brain and spinal cord, which are caused by the infiltration of immune cells (e.g., T cells and B cells).<sup>2</sup> Clonal expansion of B cells in MS CNS usually results in intrathecal antibody production.<sup>3</sup> The clinical presentation and course of MS are heterogeneous.<sup>4</sup> Symptoms of MS include mobility restrictions (oftentimes due to paresis and spasticity), visual impairment, bladder and bowel problems, pain, depression, cognitive deficits and fatigue.<sup>5</sup> Different courses of MS are distinguished<sup>6</sup>: relapsing-remitting MS (RRMS), secondary progressive MS (SPMS) and primary progressive MS (PPMS). In the early phase of the disease, 85–90% of the patients have RRMS with reversible episodes of neurological deficits (known as relapses) lasting several days or weeks. Over time, RRMS patients tend to stop having relapses but continue to develop permanent neurological deficits, which is referred to as the SPMS stage.<sup>7</sup> A minority of patients (10–15%) have PPMS, with progression of disability without acute relapses from disease onset. RRMS is typically diagnosed in young adults between 20 years and 35 years of age, whereas PPMS usually begins at ~40 years of age.<sup>1,8</sup> Moreover, it has been shown that the female-to-male ratio is 2.5:1 for patients with RRMS/SPMS and 1.2:1 for patients with PPMS.<sup>9</sup>

There is still no curative therapy for MS, but a better understanding of the immunological and neurobiological processes underlying MS has led to the development of treatments that can significantly reduce disease activity.<sup>1</sup> For instance, monthly infusions of the monoclonal antibody natalizumab, which inhibits the transendothelial migration of lymphocytes into the CNS by blocking their attachment to cell adhesion proteins, proved highly effective in preventing relapses and slowing progression of disability.<sup>10,11</sup> The newer so-called immune reconstitution therapies (IRTs) even have the potential to induce long-term drug-free remission in many patients.<sup>12</sup> With IRTs, components of the adaptive

immune system are depleted to allow for renewal. Typical representatives of IRTs are alemtuzumab and cladribine, which are administered in annual treatment courses.<sup>13</sup> A common mechanism of action is the preferential depletion of memory B cells.<sup>14</sup> Following depletion, the circulating CD19<sup>+</sup> B cells repopulate within 6–9 months [14–16]. B cells are key players as they mediate cytokine production, antigen presentation, antibody synthesis and the formation of ectopic follicles at sites of inflammation.<sup>3</sup> However, the reader is referred to the literature for a more comprehensive overview on the broad spectrum of disease-modifying treatments available for MS<sup>17–19</sup> and on the roles of B cells in MS.<sup>20,21</sup>

The exact causes of MS are still unknown, but it is well established that the individual risk of developing MS is determined by interactions of genetic and non-genetic factors. Environmental and lifestyle factors that have been found to be associated with MS risk include Epstein-Barr virus (EBV) infection, cigarette smoking, adolescent obesity and vitamin D deficiency.<sup>22,23</sup> The strongest genetic susceptibility variants for MS are located within the human leukocyte antigen (HLA) region. In particular, the *HLA-DRB1\*15:01* allele confers an increased risk of MS with an odds ratio (OR) of 3.92, whereas the *HLA-A\*02:01* allele has protective effects (OR=0.67).<sup>24</sup> In addition, more than 200 non-HLA MS risk loci with ORs ranging from 1.06 to 2.06 have been identified in large-scale genetic studies with more than 100,000 subjects of primarily European ancestry.<sup>25,26</sup> Cell type-specific enrichments of MS genetic signals have pointed to prominent roles of microglia<sup>26</sup> as well as T<sub>H</sub>17 cells and memory B cells<sup>27</sup> in the pathogenesis of MS. However, the disease-causative variants remain largely unclear, as different single-nucleotide polymorphisms (SNPs) in linkage disequilibrium (LD) can explain the association with MS.<sup>28</sup> Most disease-associated SNPs are expected to capture regulatory effects, which means that they are colocalised with quantitative trait loci (QTLs), e.g., for gene expression (eQTLs) and splicing (sQTLs).<sup>29</sup> In contrast to extensive investigations of mRNA eQTLs and sQTLs across many human tissues and cell types,<sup>30–32</sup> there are still relatively few studies on microRNA (miRNA) expression QTLs (miR-eQTLs).<sup>33,34</sup>

MicroRNAs are small non-coding RNAs about 22 nt in length that act as post-transcriptional regulators of gene expression.<sup>35</sup> A total of 2,883 human mature miRNAs are currently listed in the miRBase database release 22.<sup>36</sup> Approximately half of all known miRNAs originate from intragenic regions.<sup>37</sup> However, the expression of miRNA and host gene can be uncoupled.<sup>38</sup> Mature miRNAs are processed from stem-loop regions of longer primary miRNA (pri-miRNA) transcripts in distinct maturation steps<sup>39</sup>: A key step is the cleavage of the miRNA stem-loop by Microprocessor,<sup>40,41</sup> a nuclear complex of the RNase Drosha and its essential cofactor DGCR8. This leads to a small hairpin-shaped precursor

miRNA, which is exported into the cytoplasm where it is cleaved by the RNase Dicer. One strand of the resulting miRNA duplex is loaded into the so-called RNA-induced silencing complex, which mediates the decay and/or translational repression of target transcripts that contain a binding site for the mature miRNA. Alternative cleavage sites and post-maturation sequence modifications can give rise to isoforms of miRNAs (isomiRs) that differ in sequence from the canonical mature miRNA.<sup>42</sup>

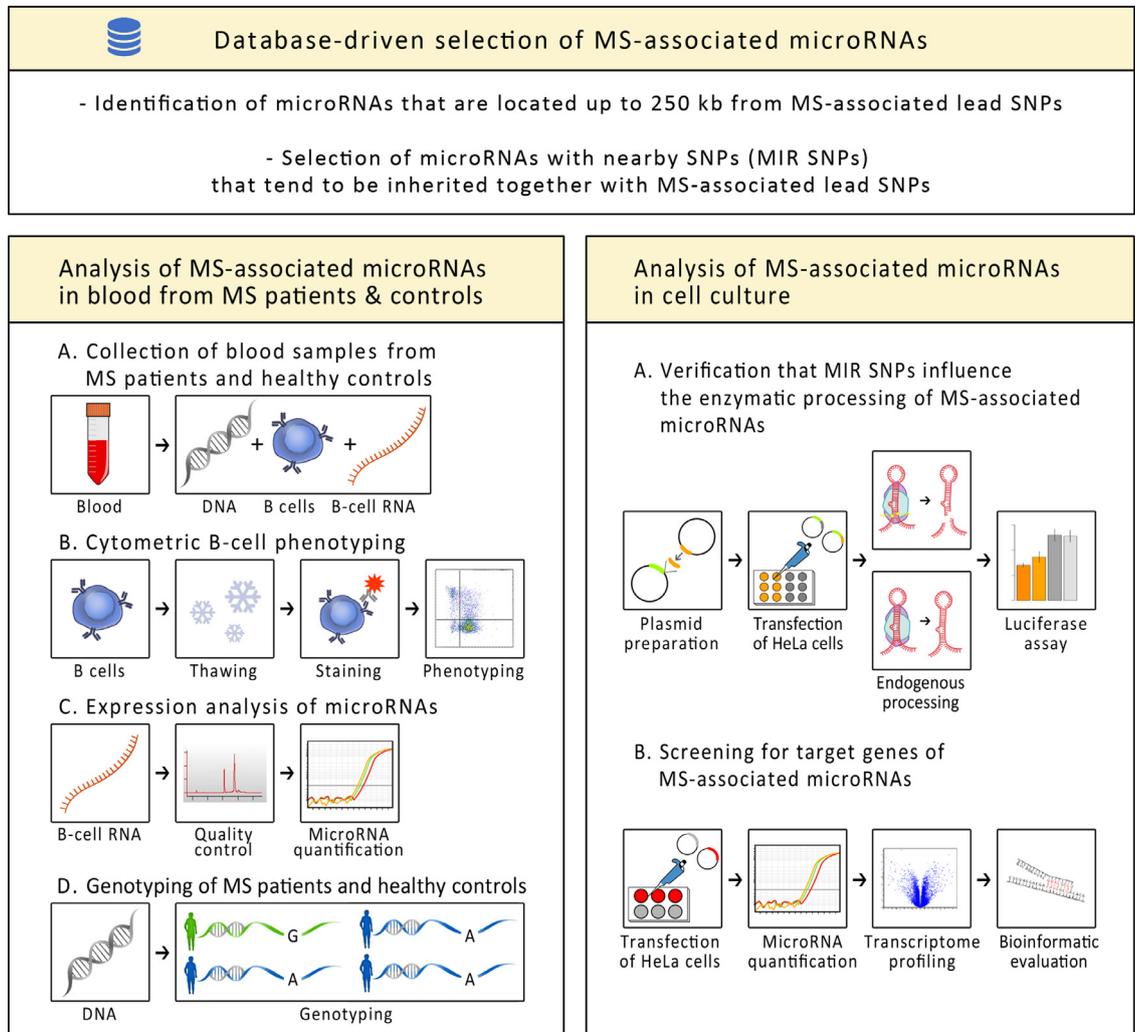
Sequence motifs in the miRNA stem-loop region are known to influence the processing by Microprocessor. These are a UG motif at position ~-14/-13 from the 5' Drosha cleavage site, a GUG/UGU motif in the apical loop and a ~16–21 nt downstream CNNC motif,<sup>43</sup> which is bound by the proteins DDX17 and SRSF3.<sup>44,45</sup> Roughly 80% of all conserved human miRNA sites contain at least one of these motifs.<sup>44</sup> Further determinants of an efficient miRNA processing are a GNNU motif and a mismatched GHG motif in the basal stem region and an extensive base pairing of the ~35-bp stem.<sup>41,46</sup> However, yet unknown sequence motifs and accessory proteins are thought to contribute to the Drosha-mediated cleavage.<sup>47</sup> SNPs that are located in features required for the recognition of pri-miRNAs can affect their processing. First studies could show a dependence of the expression of miRNAs on the genotype of MS-associated SNPs in *cis*: For carriers of the risk alleles of the SNPs rs2910164 (C) and rs1414273 (C), an increased expression of hsa-miR-146a-5p and hsa-miR-548ac-3p, respectively, was described in blood cells.<sup>48–50</sup>

In this study, we conducted a more systematic investigation of miRNAs derived from genetic loci associated with MS risk. To this end, we measured the levels of prioritised mature miRNAs and isomiRs in B cells from MS patients (n=63) and healthy controls (n=28) and analysed whether their expression is subject to *cis*-miR-eQTL effects. We then used luciferase reporter assays to verify whether the enzymatic processing of selected miRNA stem-loops is indeed causally related to the genotype of the proximal SNPs. Moreover, we aimed to identify new target genes that are post-transcriptionally regulated by the miRNAs under scrutiny. Our study provides insights into the molecular mechanisms underlying the pathogenesis of MS by uncovering the putative functional role of genetic variants implicated in disease susceptibility.

## Methods

### Selection of microRNAs from MS-associated genetic regions

A database-driven approach was employed to prioritise MS-associated genetic variants in potential pri-miRNA processing sites (Figure 1). First, we identified all human miRNA stem-loops that are recorded in the



**Figure 1.** Overview of the study.

This study comprised three parts: First, primary miRNAs encoded in genetic risk loci for MS were identified by a database-driven approach. Second, DNA, B cells and B-cell RNA were collected from MS patients as well as from healthy controls. Flow cytometry was used to phenotype B-cell subpopulations. The expression levels of mature miRNAs and isomiRs were measured using specific qPCR assays. The subjects were genotyped with respect to MS-associated SNPs within the miRNA-coding regions. Third, the effect of the SNPs on the enzymatic cleavage of the primary miRNAs by the Microprocessor complex was investigated using luciferase reporter assays. Transcriptome profiling was used to identify miRNA target genes. MIR=microRNA, MS=multiple sclerosis, qPCR=quantitative polymerase chain reaction, SNP=single-nucleotide polymorphism.

miRBase database (release 21)<sup>36</sup> and that are encoded within a distance of <250 kb to the 233 MS-associated lead SNPs, which have been reported in the most recent genome-wide association study (GWAS) in MS.<sup>26</sup> Second, we used the LDproxy and LDpair modules of the web-based tool LDlink (release 3.0)<sup>51</sup> to check whether the miRNA loci contain SNPs (hereinafter referred to as MIR SNPs) that are in LD with the MS lead SNPs. More specifically, we excluded miRNA-encoding regions for which all proximal SNPs had an  $r^2 < 0.1$  in the European subpopulation of the 1000 Genomes panel,<sup>52</sup> and we selected genetic variants with  $D' > 0.7$  that are located

within a miRBase-annotated stem-loop sequence or 50 bp upstream or downstream of the putative Drosha cleavage sites and that had a minor allele frequency (MAF) >1% according to dbSNP version 150.<sup>53</sup> If the MIR SNPs were not included in the LDlink tool, adjacent SNPs were considered as proxy. Finally, we excluded SNPs that presumably do not affect the canonical pri-miRNA processing by Drosha (e.g., SNPs in the vicinity of mirtrons<sup>39</sup>).

The prioritised pri-miRNAs may give rise to different miRNA isoforms that differ at the 5' or 3' end from the miRBase-annotated mature miRNA forms. Mature

miRNAs may also originate from both arms of a precursor miRNA, even if not explicitly recorded in the miRBase database.<sup>56</sup> Therefore, we extracted potential isomiRs from the IsomiR Bank database<sup>54</sup> and ordered them by the cumulative reads per million (RPM) values. Additionally, we inspected the normalised miRNA expression data pileup plots provided in the miRCarta database version 1.1.<sup>55</sup> Based on this information, we derived up to 2 potential isomiRs with a length of >16 bases for each arm of each precursor miRNA. The isomiRs were named following the nomenclature by Cloonan et al.<sup>56</sup>

The web-based tools RNAfold<sup>57</sup> and forna<sup>58</sup> were used to visualise the predicted minimum free energy (mfe) secondary structures of selected miRNA stem-loop sequences.

### Study groups and samples

We were interested in examining whether the selected miRNAs are dysregulated in B cells of patients with MS and whether their levels are related to the genotype of MS-associated SNPs. For this purpose, 20 ml of peripheral blood were taken from MS patients during regular clinical visits as well as from healthy donors at the Rostock University Medical Center between July 2015 and December 2019. Individuals younger than 18 years of age were excluded from this observational study. The diagnosis of MS has been confirmed according to the revised McDonald criteria.<sup>59</sup> Neurological impairment was rated using the Expanded Disability Status Scale (EDSS).<sup>60</sup> Routine medical care was provided to all patients. Hence, the patients were treated and monitored according to the guidelines and recommendations of the German Society of Neurology.

The blood samples were collected from a peripheral vein into tubes with ethylenediaminetetraacetic acid (EDTA) as anticoagulant. Immediately following the blood withdrawal, each sample was processed to obtain DNA and B cells (Figure 1). DNA was purified from 200 µl of whole blood using the QIAamp DNA Blood Mini Kit (Qiagen) and stored at -20°C. Peripheral blood mononuclear cells (PBMC) were isolated using Ficoll density gradient separation (Histopaque-1077, Sigma-Aldrich). Untouched B cells were then enriched from the PBMC by negative selection using the Pan B Cell Isolation Kit (Miltenyi Biotec). Accordingly, non-B cells were magnetically labelled with a cocktail of biotin-conjugated monoclonal antibodies and anti-biotin MicroBeads and then removed by magnetic separation following manufacturer's instructions. The number of isolated B cells was counted under a microscope using Neubauer chambers. From each sample, two aliquots of 100,000 B cells were cryopreserved in freezing medium (Biofreeze, Biochrom) in the vapor phase of liquid nitrogen at below -140°C. From the remaining B cells, total RNA was preserved using the miRNeasy Mini Kit (Qiagen) in combination with the RNase-free DNase set (Qiagen). The RNA was stored at -80°C.

### Flow cytometry

The purities of freshly isolated B cells were assessed by flow cytometry using BD FACSCalibur and BD FACSAria instruments and CD19 or CD20 antibodies conjugated to fluorescein isothiocyanate (FITC) or phycoerythrin (PE) (RRIDs: AB\_2726271, AB\_2726142 and AB\_2656064) (Miltenyi Biotec). The acquired data were analysed using Flowing Software version 2.5.1 (Turku Bioscience).

To characterise the samples in more depth, we aimed to acquire information on the composition of the peripheral B cells from the healthy subjects and the patients with MS. The B-cell phenotyping was done with the cryopreserved B cells using a BD FACSAria Illu system. For this purpose, the B cells were quickly thawed in a 37°C water bath, washed with phosphate buffered saline and then stained with the fluorochrome-conjugated antibodies CD19-PerCP (RRID: AB\_893276), CD23-BV421 (RRID: AB\_2562460), CD27-PE (RRID: AB\_314299), CD38-PE/Dazzle594 (RRID: AB\_2564104) (all from BioLegend), CD21-APC (RRID: AB\_2727123), CD24-APC-Vio770 (RRID: AB\_2656566) (Miltenyi Biotec) and IgD-BV750 (RRID: AB\_2868411) (BD). Zombie Green dye was used for live/dead staining of the cells (BioLegend). FcR blocking reagent (Miltenyi Biotec) was used to block unspecific binding of antibodies to Fc receptors. In each staining batch, 2 million cells of pooled PBMC from healthy subjects were used as an internal control. Data acquisition and compensation calculations were performed using the BD FACSDiva software version 8.0.2.

The data were further processed in the FlowJo software version 10 from BD. The FlowJo plugin CytoNorm version 0.9<sup>61</sup> was applied for normalising batch effects based on the data of the internal controls. The plugin FlowAI version 2.1<sup>62</sup> was applied for checking the data quality (e.g., the steadiness of the flow rate) and for detecting and removing outlier events. The gating strategy for the identification of B cell subsets was adapted from Morbach et al.<sup>63</sup> First, live single CD19<sup>+</sup> cells were determined based on their forward and side scatter properties and the signals for the live/dead and CD19 markers. We then gated on CD27<sup>-</sup>IgD<sup>+</sup> naive B cells, CD27<sup>+</sup>IgD<sup>+</sup> non-switched memory B cells, CD27<sup>+</sup>IgD<sup>-</sup> switched memory B cells, CD27<sup>-</sup>IgD<sup>-</sup> memory B cells, CD24<sup>++</sup>CD38<sup>++</sup> transitional B cells, CD24<sup>-</sup>CD38<sup>++</sup> plasmablasts, CD21<sup>-/low</sup>CD38<sup>-/low</sup> B cells and CD23<sup>+</sup> B cells,<sup>64</sup> as shown in Figure S1. The number of gated events was used to determine the respective B-cell subpopulation frequencies.

### Measurement of mature microRNAs in B cells

The integrity of the B-cell RNA samples was checked using an Agilent Bioanalyzer 2100 with RNA 6000 Pico kits (Agilent Technologies). RNA concentrations were measured using a NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific).

After the RNA quality check, the levels of mature miRNAs and miRNA isoforms were measured by quantitative polymerase chain reaction (qPCR) using custom Specialty TaqMan Array Cards (48 format) for standard TaqMan miRNA assays with FAM-labelled probes (Thermo Fisher Scientific). Reverse transcription (RT) was performed using 35 ng of RNA per sample and a custom target-specific stem-loop RT primer pool. The cDNA was then preamplified in 12 PCR cycles with custom PreAmp primers following the manufacturer's protocol for the use of Megaplex primer pools (Applied Biosystems). Finally, the qPCR measurements were prepared using TaqMan Universal Master Mix II with UNG and performed using a ViiA 7 Real-Time PCR System (Applied Biosystems) with 45 amplification cycles. Each miRNA was measured either in single wells or in technical triplicates, and hsa-miR-191-5p was included in the analysis as reference miRNA.<sup>65</sup>

Raw  $C_T$  values were determined using the ExpressionSuite software version 1.3 (Thermo Fisher Scientific) and averaged over replicates. Assays for which the automatic threshold algorithm failed were excluded from the further analyses. Missing values, which resulted when no  $C_T$  value could be determined within 45 PCR cycles, were imputed using the expectation-maximisation algorithm that is implemented in the R package *nondetects*.<sup>66</sup> For sensitivity analyses, the imputed dataset was truncated to a maximum  $C_T$  value of 35. The data were normalised to the reference miRNA ( $\Delta C_T$  method) and converted to the linear scale using the equation  $2^{-\Delta C_T} \times 1,000$ .

### Genotyping

The DNA extracted from the blood samples was used to determine the genotypes of MIR SNPs and tagging SNPs for *HLA-A\*02:01*<sup>67</sup> and *HLA-DRB1\*15:01*<sup>68</sup> for each subject. The genotyping was done by PCR with allele-specific TaqMan probes labelled with FAM or VIC on custom Specialty TaqMan Array Cards (Thermo Fisher Scientific). For each reaction, we used ~2 ng of DNA that was diluted in TaqMan Universal Master Mix II with UNG (Applied Biosystems). PCR amplification was carried out following manufacturer's instructions in a ViiA 7 Real-Time PCR System (Applied Biosystems). Genotype calling was performed by allelic discrimination with the TaqMan Genotyper Software version 1.6 (Applied Biosystems).

### Evaluation of primary microRNA processing efficiencies

We employed a luciferase reporter system to evaluate whether the pri-miRNA processing efficiency depends on the genotype of MS-associated SNPs (Figure 1). The assay has been adapted from Allegra et al.<sup>69,70</sup> and is based on a luciferase reporter gene that is fused with a miRNA precursor sequence. We first tested the assay

with the pri-miRNA hsa-mir-16-1 because its processing has already been extensively studied.<sup>44,71,72</sup>

We utilised the luciferase reporter vector CmiT000001-MT05 (GeneCopoeia), which encodes the secreted *Gaussia* luciferase (GLuc) as reporter gene and the secreted alkaline phosphatase (SEAP) as internal control for normalisation. In addition, we purchased miExpress precursor miRNA expression plasmids for the prioritised miRNAs as well as for hsa-mir-16-1 from GeneCopoeia. The plasmids were amplified in *E. coli* cultures and purified using the Qiagen EndoFree Plasmid Maxi Kit and QIAfilter Midi Cartridges.

Mutations resembling the naturally occurring genetic variants in the miRNA-coding sequence region (i.e., MIR SNPs) were introduced into the precursor miRNA plasmids. In the hsa-mir-16-1 precursor plasmid, an artificial C>T mutation at position 3' (+19) was introduced. This mutation has been previously shown to suppress the cleavage of hsa-mir-16-1 by the Microprocessor complex.<sup>44,73</sup> The primers for the desired point mutations were designed using the QuikChange Primer Design program from Agilent Technologies and purchased from Thermo Fisher Scientific. The mutageneses were performed using the QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies) according to the manufacturer's instructions. The success of each mutagenesis was confirmed by commercial Sanger sequencing using the EGFP-C-for primer (Microsynth Seqlab).

The different miRNA precursor sequences with the respective alleles were cloned into the luciferase reporter vector at the 3' untranslated region (UTR) of the GLuc gene. This was done by conventional restriction digestion and ligation. *Sfa*AI and either *Eco*RI or *Xho*I (Thermo Fisher Scientific) were used as restriction enzymes. The precursor miRNA fragments were prepared with the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel) and ligated into the restriction sites of the luciferase reporter vector with T4 DNA ligase (Thermo Fisher Scientific). Successful cloning was confirmed by gel electrophoresis or DNA sequencing (Microsynth Seqlab). The cloned constructs were then amplified in *E. coli* and purified using the Nucleobond Xtra Midi EF Kit (Macherey-Nagel).

HeLa cells were transiently transfected with the reporter constructs in 3 biological replicates using the FuGENE HD Transfection Reagent (Promega). The HeLa cells were originally obtained from the German Cancer Research Center (Heidelberg, Germany). They were regularly tested for mycoplasma contamination using the MycoAlert Mycoplasma Detection Kit (Lonza). For establishing the assay with the hsa-mir-16-1 constructs, four different amounts of plasmid DNA (100 ng, 200 ng, 300 ng and 500 ng) were used, and supernatants were collected at 24 h, 48 h and 72 h. Otherwise, 200 ng and 300 ng of plasmid DNA were used, and the transfected HeLa cells were cultured for 24 h

and 48 h. In this time, the stem-loop structure that is formed by the GLuc-precursor miRNA hybrid transcript can be recognised by the endogenous Microprocessor complex. A more efficient cleavage of the miRNA stem-loop leads to a preferential degradation of the luciferase transcript and thus reduced reporter protein levels. The GLuc and SEAP activities were measured in technical duplicates using the Secrete-Pair Dual Luminescence Assay Kit (GeneCopoeia) and a GloMax-Multi Detection System (Promega) according to the manufacturers' recommendations.

Afterwards, average ratios of the luminescence catalysed by GLuc and SEAP were calculated. A higher GLuc/SEAP ratio indicates a lower Droscha cleavage activity. The ratios were further normalised to those resulting for the negative control, that is, the original luciferase reporter vector without an inserted miRNA precursor fragment. The decrease in relative luminescence as compared to the negative control is a measure of the pri-miRNA processing efficiency.<sup>79</sup> In this calculation, the processing efficiency for the negative control vector is set to zero.

### Target gene analysis

We aimed to identify direct target genes of those miRNAs that were found to be processed dependent on the genotype of MS-associated SNPs. For this purpose, we conducted a transcriptome profiling analysis of HeLa cell cultures (Figure 1). The cells were transiently transfected with the respective miExpress precursor miRNA expression plasmids (GeneCopoeia) or the scrambled control clone for the pEZX-MR04 vector (CmiR0001-MR04, GeneCopoeia) as negative control using the FuGENE HD Transfection Reagent (Promega). Total RNA was isolated 24 h (from 2 replicates) and 48 h (from 4 replicates) following transfection using the miRNeasy Mini Kit (Qiagen). RNA concentrations were measured using a NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific), and RNA integrity was assessed with an Agilent 2100 Bioanalyzer using RNA 6000 Nano kits (Agilent Technologies).

To confirm the overexpression of mature miRNA molecules, a qPCR analysis was performed with pre-designed TaqMan single-tube assays from Thermo Fisher Scientific. The cDNA was generated from 10 ng intact RNA with miRNA-specific stem-loop primers and mixed with TaqMan Universal Master Mix II with UNG (Applied Biosystems). The qPCR measurement was then run in triplicates for 45 cycles on a ViiA 7 Real-Time PCR System (Applied Biosystems). As reference miRNA, hsa-miR-191-5p was chosen.<sup>65</sup> The Expression-Suite software version 1.3 (Thermo Fisher Scientific) was used to determine  $C_T$  values based on the FAM reporter signals. The average  $C_T$  values per triplicate were normalised to those of the reference miRNA to obtain  $\Delta C_T$  values.

The same RNA samples were used for the transcriptome analysis. From each sample, 210 ng of total RNA was used as starting material to generate amplified, fragmented and biotinylated single-stranded sense strand DNA using the GeneChip WT PLUS Reagent Kit (Applied Biosystems). The hybridisation on high-resolution Clariom D arrays for human (Applied Biosystems) was then carried out for 16 h at 45°C in a GeneChip Hybridization Oven 645 (Affymetrix) following the manufacturer's protocol. After washing and staining in a GeneChip Fluidics Station 450 (Affymetrix), the microarrays were scanned using a GeneChip Scanner 3000 7G (Affymetrix). The scans were imported to the Affymetrix GeneChip Command Console version 4.0 to extract the signal intensities for the >6.7 million 25mer oligonucleotide probes per array. The Transcriptome Analysis Console (TAC) version 4.0.2 (Applied Biosystems) was used for initial quality control assessments and for processing the data with the signal space transformation robust multi-array average algorithm for background adjustment, quantile normalisation, probe set summarisation and log2 transformation.

Based on the processed data, an analysis of differential gene expression was performed in TAC. Potential target genes of the miRNAs were identified by filtering those transcript clusters (TC) that were significantly lower expressed at both time points (*i.e.*, at 24 h and at 48 h), with at least 33.3% lower levels at one time point, in HeLa cells transfected with the precursor miRNA plasmids compared to HeLa cells transfected with the negative control plasmid. A nominal significance level of  $\alpha=5\%$  was chosen. The filtering result was then further confined to protein-coding genes with an official gene symbol according to the HGNC database.<sup>74</sup> Finally, to obtain the most likely direct target genes, we used RNAhybrid version 2.2<sup>75</sup> for checking whether there are predicted binding sites for the canonical miRNAs or isomiRs in the 3' UTRs of the mRNAs. The 3' UTR sequences were retrieved with the BioMart tool from Ensembl release 101,<sup>76</sup> and we requested an mfe of less than  $-16$  kcal/mol for the miRNA-target interaction as well as a perfect match with the seed region of the miRNA (*i.e.*, position 2 to 7).

We evaluated whether the genes were already listed as predicted or experimentally determined target genes in the miRWalk 3.0 database<sup>77</sup> and in miRTarBase release 8.0,<sup>78</sup> respectively. The search in miRWalk was restricted to 3' UTR miRNA binding sites. Moreover, to assign the putative target genes to functional categories, we searched them in signalling pathways using Reactome release 75,<sup>79</sup> in Gene Ontology sets using Enrichr<sup>80</sup> and in the UniProt Knowledgebase.<sup>81</sup> The categories with the greatest gene set overlap were recorded. We also studied whether or not the genes are expressed in cell types of the blood and brain using the microarray data from Novershtern et al.<sup>82</sup> (GEO accession

GSE24759) and the RNA-sequencing data from Darnis et al.<sup>83</sup> and by demanding a transcript level of >100.

### Statistical analyses

The statistical analyses were performed in R version 3.6.2 unless stated otherwise. For descriptive statistics, means and standard deviations (SD) per group were calculated. The inferential statistics were exploratory in nature. Therefore, the significance level was generally set at  $\alpha=0.05$  to minimise the risk of type II errors. The data were visualised in beeswarm/violin plots and bar plots using the R packages beeswarm, vioplot and gplots, respectively.

We tested for differences in B-cell subpopulation frequencies and miRNA expression levels (in linear scale) between the study groups using linear models while adjusting for age and sex as potential confounding variables. Type II *F*-tests were calculated for all main effects in the models with the R package car.<sup>84</sup> Pairwise comparisons between the study groups were performed with Tukey post hoc tests adjusted for age and sex using the R package multcomp.<sup>85</sup> We used Pearson correlation tests to compare the levels of canonical miRNAs and their potential isomiRs.

The SNP genotype distributions were compared between the MS patients and healthy controls using one-tailed Fisher's exact tests in order to evaluate the association of risk alleles with MS in our study population. Additionally, allelic ORs were calculated using the Wald method. The effect of each MIR SNP on the expression level of the respective mature miRNA was examined in *cis*-miR-eQTL analyses. For this purpose, we fitted linear mixed-effects models using the R package lme4.<sup>86</sup> The study group was treated as a random effect. Genotype, age and sex were considered as fixed effects. The models were fitted without interaction term and with the genotype effect being captured in an additive fashion using the restricted maximum likelihood approach. Significance values were then computed with type II Wald  $\chi^2$  tests using the R package car.<sup>84</sup> For pairwise genotype comparisons, Tukey post hoc tests adjusted for age, sex and group were carried out.

The balanced dataset from the luciferase reporter assays was analysed using an additive 3-way analysis of variance (ANOVA), with amount of plasmid DNA, time point and genotype as independent variables and GLuc/SEAP ratios or the decrease in relative luminescence as dependent variable. Accordingly, *F*-tests were used to examine the significance of individual effects. Welch two-tailed *t*-tests were applied for comparing the data that were obtained under the same experimental conditions with plasmids, which represent the two alleles of a particular SNP.

### Ethics

This study was conducted with approvals by the local ethics committee of the University of Rostock (permit

numbers A 2014-0112 and A 2016-0188) and in compliance with the principles of the Declaration of Helsinki. All subjects gave prior written informed consent for the scientific use of their biosamples.

### Role of funders

The funders had no role in the study design, data collection, analysis and interpretation, decision to publish or preparation of the manuscript.

## Results

### Prioritised microRNAs from genetic risk loci for MS

We identified a total of 11 pri-miRNAs from MS-associated genetic regions that were considered for the further investigations in the present study (Table 1). These pri-miRNAs contain SNPs in the adjacent 50 bp around the miRNA precursor sequence that were suspected to influence the Drosha-mediated enzymatic cleavage of the stem-loop. In case of hsa-mir-4304, two such MIR SNPs were found. Seven of the 12 ascertained MIR SNPs are in complete LD ( $D'=1$ ) with the nearby MS lead SNP from the GWAS,<sup>26</sup> which means that one MIR SNP allele is always inherited together with one specific MS lead SNP allele according to the 1000 Genomes haplotype data.<sup>52</sup> Six of the 12 MIR SNPs are rare genetic variants with a global MAF of less than 2% (Table 1).

Except for hsa-mir-26a-2 and hsa-mir-199a-1, all selected pri-miRNAs are primate-specific,<sup>36</sup> and all were classified as Drosha-dependent in the study by Kim et al.<sup>87</sup> Six of the miRNA stem-loops are presumably processed from introns of host gene transcripts (Table 1). Our research group has previously explored the role of hsa-mir-548ac, which is encoded in the first intron of the *CD58* gene, and we found that MS risk allele carriers have increased levels of the mature miRNA in blood-derived cells.<sup>48,49</sup> Another prioritised pri-miRNA is hsa-mir-934, which is located in the fourth intron of the *VGLL1* gene, the only sex chromosomal MS risk locus that has been reported in the GWAS from 2019.<sup>26</sup> The majority of the selected miRNAs ( $n=7$ ) have a 4-digit number. This means that they were discovered relatively recently, that they are usually not covered by standard qPCR- or array-based miRNA measurement platforms, and that their regulatory functions are often still largely unclear.

Three of the MIR SNPs affect the basal stem (Table 1). For instance, the A allele of SNP rs73558572 causes a mismatch in the hsa-mir-934 secondary structure. All the other MIR SNPs are located in the flanking segments. Three of them belong to a downstream CNNC motif (rs817478, rs1005039 and rs41292017). We hypothesised that the pri-miRNA processing

Primary microRNA	miRBase accession	Mature forms	Chromosome (strand)	Gene locus	MS lead SNP identifier <sup>a</sup>	MIR SNP identifier	SNP distance in bp <sup>b</sup>	LD D' (EUR)	MIR SNP allele frequency <sup>b</sup>	MIR SNP position	MIR SNP location
hsa-mir-26a-2	MI0000750	5p and 3p	chr12 (-)	<i>CTDSP2</i>	rs701006	rs41292017	111546	1.0000	G: 98.2%; A: 1.8%	3' (+21)	CNNC
hsa-mir-199a-1	MI0000242	5p and 3p	chr19 (-)		rs12609500	rs1005039	245847	1.0000	C: 98.6%; T: 1.4%	3' (+24)	CNNC
hsa-mir-548ac	MI0016762	3p	chr1 (-)	<i>CD58</i>	rs10801908	rs1414273	12156	0.9916	C: 57.2%; T: 42.8%	3' (+11)	basal stem
hsa-mir-934	MI0005756	5p	chrX (+)	<i>VGLL1</i>	rs2807267	rs73558572	33408	0.9674 <sup>c</sup>	G: 99.0%; A: 1.0%	5' (-6)	basal stem
hsa-mir-3661	MI0016062	5p	chr5 (+)		rs244656	rs779704489	111732	0.7328 <sup>d</sup>	A: 98.1%; C: 1.9%	3' (+34)	flanking region
hsa-mir-3671	MI0016072	3p	chr1 (-)	<i>JAK1</i>	rs72922276	rs521188	94200	1.0000	A: 89.3%; G: 10.7%	5' (-6)	basal stem
hsa-mir-4252	MI0015864	3p	chr1 (-)		rs2986736	rs2003071	22681	0.7869	C: 91.1%; T: 8.9%	3' (+38)	flanking region
hsa-mir-4304	MI0015832	5p	chr12 (-)	<i>PITPNM2</i>	rs7975763	rs78351440, rs77905262	108764, 108773	1.0000, 1.0000	C: 94.0%; A: 6.0%, G: 89.3%; C: 10.7%	5' (-24), 5' (-15)	flanking region, flanking region
hsa-mir-4423	MI0016760	5p and 3p	chr1 (+)		rs35486093	rs817478	130257	0.7445	A: 73.2%; C: 26.8%	3' (+18)	CNNC
hsa-mir-4464	MI0016812	5p	chr6 (+)	<i>AL132996.1</i>	rs72928038	rs77896647	45799	1.0000 <sup>c</sup>	G: 98.9%; A: 1.1%	3' (+22)	flanking region
hsa-mir-4492	MI0016854	3p	chr11 (+)		rs12365699, rs6589706	rs7926599	38244, 33717	1.0000, 1.0000	G: 98.5%; C: 1.5%	3' (+40)	flanking region

**Table 1: Prioritised microRNAs from MS-associated genetic regions.**

Eleven miRNA-coding regions were identified in the vicinity of MS-associated SNPs. A total of 12 SNPs are located in these sequence regions and are in LD with the MS lead SNPs. These MIR SNPs were suspected to affect the Drosha-mediated processing of the miRNA stem-loop sequences. For three of the primary miRNAs, two mature miRNA molecules (from the 5' and 3' arm) are annotated in the miRBase database.<sup>36</sup> Six of the miRNA loci overlap with intronic gene regions. The position of the MIR SNPs relative to the predicted Drosha cleavage sites is indicated. Three of the MIR SNPs modify a downstream CNNC motif. bp=base pairs, EUR=European, LD=linkage disequilibrium, MIR/miRNA=miRNA, MS=multiple sclerosis, SNP=single-nucleotide polymorphism.

<sup>a</sup> taken from the latest genome-wide association study in MS.<sup>26</sup>

<sup>b</sup> distances and global frequencies of the forward strand alleles are given according to dbSNP build 151 and the GRCh38 reference genome assembly.

<sup>c</sup> the LDpair tool<sup>31</sup> did not provide LD measures for the European population, thus r2 and D' are provided for the whole 1000 Genomes panel.<sup>32</sup>

<sup>d</sup> no information for the MIR SNP in LDpair,<sup>31</sup> thus SNP rs3895755 served as proxy.

Study group	Samples, n	Subjects, n	Women, n	Men, n	Age in years, mean $\pm$ SD	Duration of disease in years, mean $\pm$ SD	Relapses in previous year, mean $\pm$ SD	EDSS score, mean $\pm$ SD (mv)
Healthy controls	28	28	17	11	28.0 $\pm$ 8.9	—	—	—
MS patients	92	63	38	25	39.4 $\pm$ 13.1	8.3 $\pm$ 6.6	0.3 $\pm$ 0.7	3.0 $\pm$ 1.6 (10)
– PPMS	13	13	5	8	58.7 $\pm$ 9.8	9.7 $\pm$ 4.6	0.0 $\pm$ 0.0	4.9 $\pm$ 1.7
– RRMS, natalizumab	29	29	16	13	44.4 $\pm$ 9.0	14.0 $\pm$ 6.1	0.0 $\pm$ 0.0	3.2 $\pm$ 1.7 (8)
– RRMS, before IRT	7	7	5	2	33.7 $\pm$ 12.2	2.8 $\pm$ 2.8	1.1 $\pm$ 1.2	2.1 $\pm$ 0.8
– RRMS, alemtuzumab	37	15	12	3	29.2 $\pm$ 4.8	4.2 $\pm$ 3.5	0.5 $\pm$ 0.7	2.4 $\pm$ 1.0 (1)
– RRMS, cladribine	6	6	5	1	42.0 $\pm$ 12.5	8.9 $\pm$ 9.0	0.5 $\pm$ 0.8	2.5 $\pm$ 0.8 (1)

**Table 2: Basic characteristics of the study groups.**

A total of 120 blood samples from 91 subjects were analysed in this study. Up to four blood samples were collected per patient. From the 7 patients that provided a blood sample immediately before the initiation of IRT, at least one sample was also obtained after the administration of alemtuzumab (n=4) or cladribine (n=3). Age and clinical characteristics are reported for the time point of blood collection. For 10 samples, no information about the patients' current degree of disability (as rated by the EDSS<sup>60</sup>) was available, and thus the means and SDs refer to cases with valid data. — = not available, EDSS=Expanded Disability Status Scale, IRT=immune reconstitution therapy, MS=multiple sclerosis, mv=missing values, n=number, PPMS=primary progressive multiple sclerosis, RRMS=relapsing-remitting multiple sclerosis, SD=standard deviation.

efficiency and/or the Drosha cleavage sites are dependent on the allele status of these SNPs.

#### Characteristics of the study population and samples

The study population comprised 91 subjects in total. From these, 120 blood samples were collected and considered eligible for the analyses in the present study. The study population was divided into 6 study groups: healthy controls (n=28), patients with PPMS who were treated with glucocorticosteroid pulse therapy (n=13) and four groups of RRMS patients (n=50) that were differentiated by the type of treatment received (Table 2). A subset of 29 RRMS patients was treated with natalizumab. Seven RRMS patients were treated with fingolimod (n=3), glatiramer acetate (n=3) or interferon- $\beta$ -1b (n=1) before switching to an IRT (RRMS before IRT group). These patients received a first treatment course of intravenous alemtuzumab (n=4) or oral cladribine (n=3) immediately after the blood collection, and they all donated at least one additional blood sample at a later time point during therapy. The samples from patients under therapy with alemtuzumab (n=37) or cladribine (n=6) were obtained at least 4 months after the last treatment week, *i.e.*, after partial B-cell repopulation following B-cell depletion. In the alemtuzumab group, one to four blood samples were collected per patient.

There was a female preponderance in the healthy control group and in the RRMS groups but not in the PPMS group (Table 2). The average age of the controls was 28.0 years. The patients were older, with a mean age of 39.4 years and a mean disease duration of 8.3 years. They had, on average, 0.3 relapses in the year prior to the blood sampling and a mean EDSS score of 3.0. The patients with PPMS had the highest age and the highest degree of disability on average. The patients who were about to switch to an IRT were in an early phase of the disease, and they had the highest relapse

rate. To account for these differences, we controlled for potential confounding by age, sex and group assignment in the statistical analyses of the data described hereafter.

The number of cells after enrichment of B cells averaged 4.3 million. Flow cytometry of the freshly isolated cells revealed a mean B-cell purity of 85.2% (2 out of 120 samples had missing values). Using the cryopreserved B cells, we were able to distinguish 8 B-cell subpopulations based on the expression of cell surface markers. The mean percentages of the different B-cell subsets across the 6 study groups are shown in Table S1. In general, CD27<sup>+</sup>IgD<sup>+</sup> naive B cells accounted for the largest proportion (up to 72.4% on average), whereas transitional and CD21<sup>−/low</sup>CD38<sup>−/low</sup> B cells occurred at the lowest frequencies (up to 7.4% and 9.0% on average, respectively). When comparing the 6 study groups, there were significant differences in the proportions of all subpopulations (*F*-test *P*<0.05). For instance, CD27<sup>+</sup>IgD<sup>+</sup> non-switched memory B cells represented ~24% of CD19<sup>+</sup> cells in the healthy controls. The respective percentages were higher, on average, in RRMS patients treated with natalizumab (43.0%) and much lower in RRMS patients on IRT with alemtuzumab or cladribine (up to 8.9% on average). Age and sex did not explain the variance in these data (*F*-test *P*>0.05).

#### Differential expression of mature microRNAs in B cells

According to miRBase,<sup>36</sup> 14 mature miRNAs are produced from the prioritised pri-miRNAs from MS-associated genetic loci (Table 1). Standard TaqMan assays were available for 13 of these. The qPCR assay design failed for hsa-miR-4492-3p due to a high GC content of 94%. To consider the possible formation of different mature miRNAs from the same stem-loop, we utilised information from IsomiR Bank<sup>54</sup> and miRCarta<sup>55</sup> to

define potential 3' isomiRs (n=7), 5' and 3' isomiRs (n=2) and putative miRNAs that may be generated from the other arm than annotated in miRBase<sup>36</sup> (n=6). Thus, a total of 28 assays were used for the measurement of those miRNAs in B cells from MS patients and healthy controls.

The data of 22 qPCR assays could be included in the further analyses. For each study group, the average expression levels relative to the reference miRNA hsa-miR-191-5p, which could be detected with  $C_T$  values <20 in all samples, are given in Table 3. For the other 6 assays, we did not obtain valid data as the automatic threshold algorithm failed (THOLDFAIL flag): hsa-miR-934-5p|[isomiR]|14\_33|, hsa-miR-3661-5p|[hsa-miR-3661]|20\_41|, hsa-miR-3661-5p|[isomiR]|20\_40|, hsa-miR-3671-5p|[isomiR]|3\_19|, hsa-miR-3671-5p|[isomiR]|3\_19|sub.6.T>C and hsa-miR-4252-3p|[hsa-miR-4252]|34\_52|.

Six miRNAs were detected in less than half of the samples: hsa-miR-548c-5p and -3p, hsa-miR-934-5p and -3p, hsa-miR-3671-3p and hsa-miR-4464-3p. The highest expression levels (mean  $\Delta C_T$  values <10) were seen for hsa-miR-26a-5p and hsa-miR-199a-3p isoforms as well as for hsa-miR-4492-3p|[isomiR]|58\_74|. In those cases where both the canonical miRNA and an isomiR from the same arm of the stem-loop were measured, the data were highly correlated (Pearson correlation  $P \leq 0.0005$ ). This was also true for the two forms of hsa-miR-4464-5p that are shifted by 4 nt. Therefore, only one form was considered for the visualisation of the qPCR data. However, it should be noted that the levels of the 3' isomiR of hsa-miR-199a-5p were much higher than those of the corresponding canonical miRNA (mean  $\Delta C_T$  11.58 vs. 20.94). Moreover, only one mature miRNA is recorded in miRBase<sup>36</sup> for hsa-mir-548c, hsa-mir-934, hsa-mir-4304 and hsa-mir-4464, but we could measure miRNA molecules derived from both the 5' arm and the 3' arm.

In comparison of the study groups, significant expression differences were found for 7 canonical miRNAs and 6 isomiRs ( $F$ -test  $P < 0.05$ ) (Table 3). These differences were also significant in the sensitivity analysis (*i.e.*, after replacing  $C_T$  values >35 by a value of 35). The differentially expressed mature miRNAs originate from 6 pri-miRNAs: hsa-mir-26a-2, hsa-mir-199a-1, hsa-mir-4304, hsa-mir-4423, hsa-mir-4464 and hsa-mir-4492 (Figure 2). Compared with the healthy controls, the levels of hsa-miR-26a-2-3p were significantly lower in RRMS patients treated with alemtuzumab (Tukey's test  $P = 0.048$ ). In contrast, the levels of hsa-miR-199a-5p and -3p were higher in alemtuzumab-treated cases than in the controls. Significantly higher levels of hsa-miR-4304-5p were measured in RRMS patients receiving cladribine in comparison to the healthy group and all other RRMS groups (Tukey's test  $P \leq 0.003$ ). In the B-cell samples of the cladribine-treated patient group, we also found, on average, the highest levels of hsa-miR-4464-

5p and a 1 nt shifted isomiR of hsa-miR-4492-3p. The average expression of hsa-miR-4423-3p|[isomiR]|48\_67| was highest in patients with PPMS.

#### Association of genetic variants with microRNA expression

The 12 MIR SNPs as well as two HLA allele tagging SNPs were genotyped for all subjects (Table S2). Overall, the allele frequencies in the study population roughly corresponded to the global allele frequencies (Table 1). We could confirm a significant association between the risk allele of SNP rs1414273 (OR=2.52) as well as the *HLA-DRB1\*15:01* allele (OR=2.69) and MS (Fisher's test  $P < 0.05$ ). However, for 4 MIR SNPs, we did not detect the rare allele in either MS patients or controls.

The *cis*-miR-eQTL analysis was conducted for hsa-mir-26a-2, hsa-mir-199a-1, hsa-mir-4304, hsa-mir-4423 and hsa-mir-4492. For these pri-miRNAs, we were able to measure corresponding mature miRNAs in at least half of the samples, on the one hand, and to detect both alleles of the MIR SNPs, on the other hand. For the genotype of SNP rs817478, we found a significant association with the levels of hsa-miR-4423-5p in B cells (Wald test  $P < 0.05$ ) (Table 4). The levels of the canonical miRNA and the respective 3' isomiR were found to be significantly lower in individuals that are homozygous for the risk allele A (Tukey's test  $P \leq 0.008$ ). A similar trend was seen for hsa-miR-4423-3p (Figure 3). We also observed higher levels of hsa-miR-199a-5p and -3p in carriers of the risk allele T of SNP rs1005039 (Figure 3). These trends did not reach statistical significance (Wald test  $P = 0.702$  and  $P = 0.158$ , respectively) (Table 4), with the analysis being hampered by the low MAF. For hsa-mir-26a-2, hsa-mir-4304 and hsa-mir-4492, we did not find evidence of a genotype-dependent processing. The results were essentially the same when using the truncated data in the sensitivity analysis.

#### Verification of allele-specific primary microRNA processing efficiencies

To confirm that MS-associated genetic variants in pri-miRNA processing sites causally affect the Drosha-mediated stem-loop cleavage, a luciferase reporter system was utilised. The assay was first established with the pri-miRNA hsa-mir-16-1. The wild-type hsa-mir-16-1 was efficiently processed across the entire range of experimental conditions, as reflected by much lower relative luminescence signals in comparison to the negative control (Figure S2). In contrast, the cleavage activity was completely lost when disrupting the CNNC motif in the flanking region by introducing a C>T mutation at position 3' (+19) (ANOVA  $P = 6.7 \times 10^{-29}$ ).

To further substantiate the results of the *cis*-miR-eQTL analysis, we then tested whether the SNPs

MicroRNA assay <sup>a</sup>	MicroRNA sequence	Detection rate <sup>b</sup> , %	Healthy controls (n=28), mean $\pm$ SD (mv <sup>c</sup> )	PPMS (n=13), mean $\pm$ SD (mv <sup>c</sup> )	RRMS, natalizumab (n=29), mean $\pm$ SD (mv <sup>c</sup> )	RRMS, before IRT (n=7), mean $\pm$ SD (mv <sup>c</sup> )	RRMS, alemtuzumab (n=37), mean $\pm$ SD (mv <sup>c</sup> )	RRMS, cladribine (n=6), mean $\pm$ SD (mv <sup>c</sup> )	P-value <sup>d</sup>
hsa-miR-26a-2-5p[hsa-miR-26a-5p]13_34	UUCAAGUAAUCCAGGAUAGGCU	100.0	2.06 $\pm$ 0.35	1.99 $\pm$ 0.55	1.76 $\pm$ 0.43	1.98 $\pm$ 0.29	2.16 $\pm$ 0.44	2.27 $\pm$ 0.27	<b>0.0014</b>
hsa-miR-26a-2-5p[iso-miR]13_33	UUCAAGUAAUCCAGGAUAGGC	100.0	0.69 $\pm$ 0.43	0.58 $\pm$ 0.67	0.36 $\pm$ 0.46	0.70 $\pm$ 0.49	0.87 $\pm$ 0.56	0.77 $\pm$ 0.08	<b>0.0011</b>
hsa-miR-26a-2-3p[hsa-miR-26a-2-3p]51_72	CCUAUUCUUGAUUACUUGUUUC	91.7	13.32 $\pm$ 1.56 (1)	13.97 $\pm$ 1.07 (5)	13.31 $\pm$ 0.99 (1)	14.33 $\pm$ 1.39	13.56 $\pm$ 0.80 (2)	13.18 $\pm$ 0.89 (1)	<b>0.0418</b>
hsa-miR-199a-1-5p[hsa-miR-199a-5p]5_27	CCCAGUGUUCAGACUACCGUUC	50.8	21.20 $\pm$ 5.30 (15)	20.76 $\pm$ 5.40 (7)	25.30 $\pm$ 5.32 (20)	21.12 $\pm$ 6.28 (4)	17.79 $\pm$ 4.58 (11)	18.28 $\pm$ 4.62 (2)	<b>0.0177</b>
hsa-miR-199a-1-5p[iso-miR]5_26	CCCAGUGUUCAGACUACCGUU	98.3	12.12 $\pm$ 2.27	12.38 $\pm$ 1.97 (1)	12.59 $\pm$ 1.64	10.42 $\pm$ 1.88	10.20 $\pm$ 2.16 (1)	12.45 $\pm$ 3.24	<b>0.0005</b>
hsa-miR-199a-1-3p[hsa-miR-199a-3p]46_67	ACAGUAGUCGACACAUUGGUUA	100.0	7.98 $\pm$ 1.54	8.12 $\pm$ 1.43	8.87 $\pm$ 1.25	6.86 $\pm$ 1.83	6.71 $\pm$ 2.03	7.72 $\pm$ 1.89	<b>0.0006</b>
hsa-miR-199a-1-3p[iso-miR]46_66	ACAGUAGUCGACACAUUGGUU	100.0	8.15 $\pm$ 1.25	8.24 $\pm$ 0.99	8.75 $\pm$ 0.88	7.43 $\pm$ 1.60	7.17 $\pm$ 1.74	8.11 $\pm$ 1.76	<b>0.0008</b>
hsa-miR-548ac-5p[iso-miR]16_36	AAAGUUAUUGUGUUUUUGCU	13.3	29.92 $\pm$ 5.06 (22)	44.07 $\pm$ 0.95 (13)	34.57 $\pm$ 4.77 (27)	31.00 $\pm$ 6.02 (6)	30.58 $\pm$ 5.79 (31)	28.28 $\pm$ 6.14 (5)	0.8165
hsa-miR-548ac-3p[hsa-miR-548ac]52_73	CAAAAACCGGCAAUUACUUUUG	30.0	25.95 $\pm$ 5.32 (21)	28.16 $\pm$ 4.72 (11)	22.94 $\pm$ 5.08 (15)	25.46 $\pm$ 5.82 (5)	25.43 $\pm$ 5.38 (26)	40.48 $\pm$ 1.20 (6)	0.8473
hsa-miR-934-5p[hsa-miR-934]14_35	UGUCUACUACUGGAGACACUGG	12.5	29.39 $\pm$ 5.61 (23)	42.23 $\pm$ 0.95 (13)	31.75 $\pm$ 5.09 (26)	29.80 $\pm$ 5.90 (6)	30.56 $\pm$ 4.46 (31)	41.49 $\pm$ 1.20 (6)	0.6109
hsa-miR-934-3p[iso-miR]50_71	GAGUCUCCAGUAAUGGACGGGA	18.3	27.58 $\pm$ 4.82 (21)	31.40 $\pm$ 4.31 (12)	30.16 $\pm$ 4.91 (25)	29.34 $\pm$ 4.65 (6)	29.21 $\pm$ 4.84 (30)	25.64 $\pm$ 4.74 (4)	0.9388
hsa-miR-3671-3p[hsa-miR-3671]59_80	AUCAAAUAAGGACUAGUCUGCA	13.3	33.75 $\pm$ 1.42 (27)	37.04 $\pm$ 0.95 (13)	31.33 $\pm$ 2.71 (27)	23.75 $\pm$ 4.79 (4)	27.03 $\pm$ 4.29 (29)	24.11 $\pm$ 4.73 (4)	0.1381
hsa-miR-4304-5p[hsa-miR-4304]10_26	CCGGCAUGUCCAGGGCA	100.0	10.72 $\pm$ 0.72	10.12 $\pm$ 0.98	10.84 $\pm$ 0.57	10.63 $\pm$ 0.81	10.92 $\pm$ 0.84	9.27 $\pm$ 1.13	<b>3.3 <math>\times</math> 10<sup>-5</sup></b>
hsa-miR-4304-3p[iso-miR]35_52	GCUCUGUGACUGCGCCA	100.0	10.74 $\pm$ 0.72	10.38 $\pm$ 1.17	11.69 $\pm$ 0.82	10.84 $\pm$ 0.98	11.01 $\pm$ 0.86	9.84 $\pm$ 0.82	<b>8.2 <math>\times</math> 10<sup>-5</sup></b>
hsa-miR-4423-5p[hsa-miR-4423-5p]12_33	AGUUGCCUUUUUGUCCCAUGC	54.2	21.19 $\pm$ 4.59 (13)	23.50 $\pm$ 4.57 (9)	19.25 $\pm$ 3.64 (7)	24.88 $\pm$ 4.68 (3)	22.35 $\pm$ 4.69 (21)	20.21 $\pm$ 4.60 (2)	0.9757
hsa-miR-4423-5p[iso-miR]12_32	AGUUGCCUUUUUGUCCCAUG	57.5	19.69 $\pm$ 3.88 (11)	21.69 $\pm$ 4.49 (8)	18.49 $\pm$ 3.34 (5)	21.76 $\pm$ 4.65 (4)	21.51 $\pm$ 3.58 (18)	25.73 $\pm$ 2.64 (5)	0.0652

Table 3 (Continued)

MicroRNA assay <sup>a</sup>	MicroRNA sequence	Detection rate <sup>b</sup> , %	Healthy controls (n=28), mean $\pm$ SD (mv <sup>c</sup> )	PPMS (n=13), mean $\pm$ SD (mv <sup>c</sup> )	RRMS, natalizumab (n=29), mean $\pm$ SD (mv <sup>c</sup> )	RRMS, before IRT (n=7), mean $\pm$ SD (mv <sup>c</sup> )	RRMS, alemtuzumab (n=37), mean $\pm$ SD (mv <sup>c</sup> )	RRMS, cladribine (n=6), mean $\pm$ SD (mv <sup>c</sup> )	P-value <sup>d</sup>
hsa-miR-4423-3p [hsa-miR-4423-3p] 48_68]	AUAGGCACCAAAAAGCAACAA	100.0	13.40 $\pm$ 1.00	13.09 $\pm$ 1.57	13.56 $\pm$ 0.85	13.57 $\pm$ 1.00	13.52 $\pm$ 0.95	12.64 $\pm$ 0.67	<b>0.0413</b>
hsa-miR-4423-3p [iso-miR] 48_67]	AUAGGCACCAAAAAGCAACA	100.0	13.03 $\pm$ 0.99	12.53 $\pm$ 1.09	13.30 $\pm$ 0.65	13.21 $\pm$ 0.97	13.14 $\pm$ 0.96	12.63 $\pm$ 0.71	<b>0.0227</b>
hsa-miR-4464-5p [hsa-miR-4464] 11_31]	AAGGUUUGGAUAGAUGCAAUA	100.0	15.91 $\pm$ 1.29	15.11 $\pm$ 1.06	16.30 $\pm$ 1.04	16.55 $\pm$ 1.36	15.88 $\pm$ 0.97	14.55 $\pm$ 1.00	<b>0.0063</b>
hsa-miR-4464-5p [iso-miR] 15_35]	UUUGGAUAGAUGCAAUAAAGU	71.7	18.66 $\pm$ 3.73 (9)	20.00 $\pm$ 4.20 (7)	18.24 $\pm$ 3.82 (7)	16.71 $\pm$ 2.52 (1)	17.52 $\pm$ 3.70 (7)	20.73 $\pm$ 3.99 (3)	0.8573
hsa-miR-4464-3p [iso-miR] 61_79]	UUAUUGCAUCUAUCCAAAC	5.8	42.23 $\pm$ 0.94 (28)	29.56 $\pm$ 5.65 (11)	32.30 $\pm$ 5.21 (26)	42.45 $\pm$ 0.81 (7)	34.71 $\pm$ 4.21 (35)	41.01 $\pm$ 1.20 (6)	0.8342
hsa-miR-4492-3p [iso-miR] 58_74]	GGGCUGGGCGCGCCCA	100.0	6.34 $\pm$ 1.26	5.85 $\pm$ 1.27	7.14 $\pm$ 1.10	7.01 $\pm$ 0.89	6.84 $\pm$ 0.98	5.24 $\pm$ 0.79	<b>0.0006</b>

**Table 3: Relative expression of mature microRNAs in B cells from MS patients and controls.**

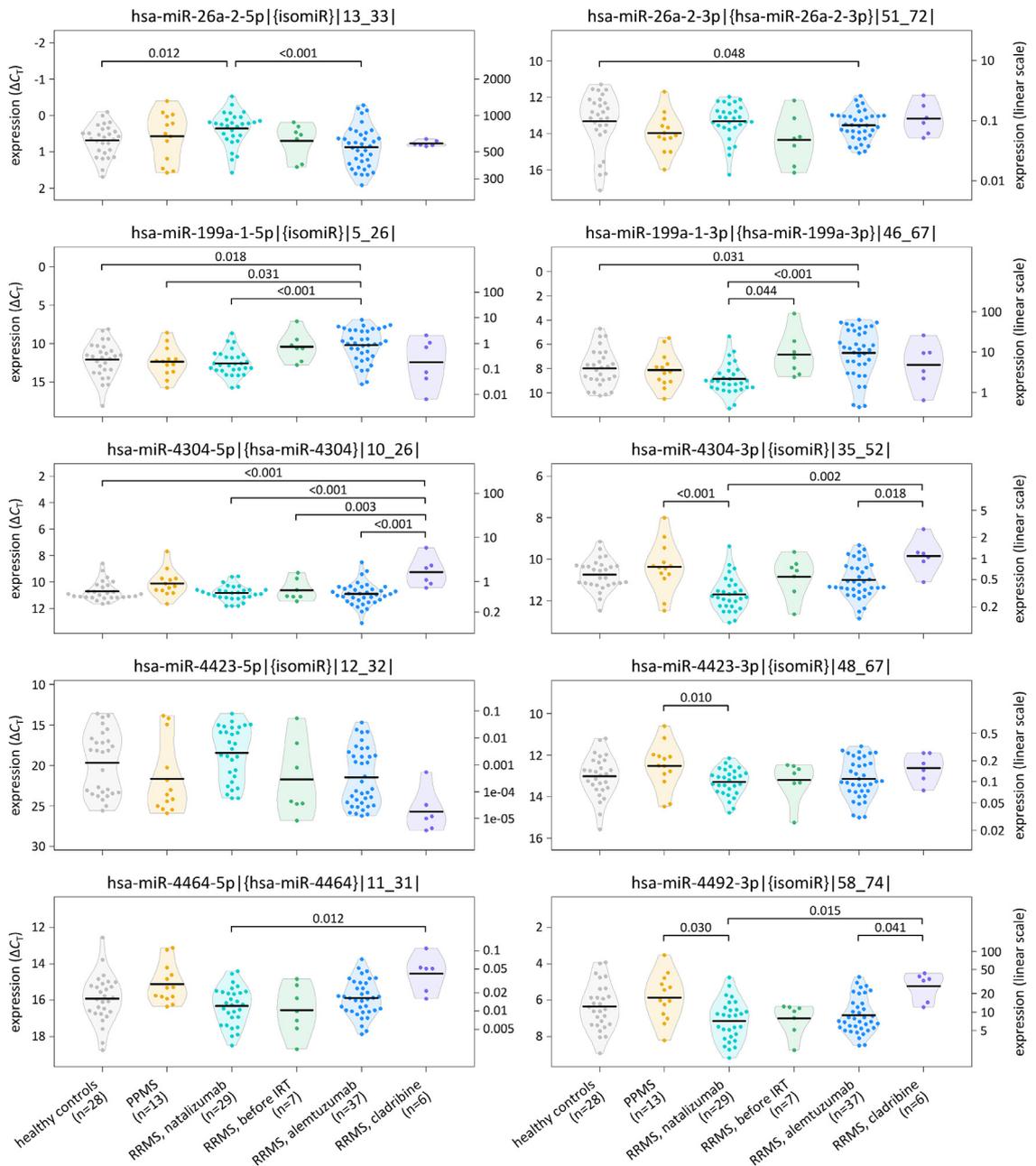
A total of 120 B-cell samples were analysed, and 22 qPCR assays gave valid data. For 6 primary miRNAs, at least one mature miRNA molecule (from the 5' or 3' arm) could be detected in at least half of the samples. Average  $\Delta C_T$  values are reported as a measure of the miRNA expression relative to hsa-miR-191-5p (reference miRNA) for each of the 6 study groups. Lower  $\Delta C_T$  values correspond to higher expression levels. The P-values indicate whether there is a difference in the level of miRNA expression between the groups. Significant differences ( $P < 0.05$ ) are marked in bold. IRT=immune reconstitution therapy, miRNA=microRNA, MS=multiple sclerosis, mv=missing values, n=number, PPMS=primary progressive multiple sclerosis, qPCR=quantitative polymerase chain reaction, RRMS=relapsing-remitting multiple sclerosis, SD=standard deviation.

<sup>a</sup> the miRNAs are specified following the nomenclature by Cloonan et al.<sup>36</sup>

<sup>b</sup> the percentage of samples for which a  $C_T$  value could be determined within 45 qPCR cycles.

<sup>c</sup> the number of samples in which the miRNA could not be detected and for which  $C_T$  values were thus imputed.

<sup>d</sup> F-test P-values for linear models fitted to the data in linear scale and adjusted for age and sex.



**Figure 2.** Differential expression of microRNAs in B cells from MS patients and healthy controls.

The expression of 10 mature miRNAs and isomiRs is visualised for the six study groups. These miRNAs could be detected in >50% of the samples and are derived from 6 primary miRNA transcripts from MS-associated genetic regions. For this plot, we always selected the 5p and 3p miRNA forms (canonical miRNA or isomiR) that had the lowest P-value in the group comparisons (Table 3). The expression was quantified relative to the reference miRNA hsa-miR-191-5p. Higher data points indicate higher expression levels. The y-axis on the left displays  $\Delta C_T$  values in an inverted manner and the y-axis on the right displays the data converted in linear scale ( $2^{-\Delta C_T} \times 1000$ ). Black horizontal lines indicate the means per group. Significance values <0.05 from pairwise Tukey post hoc tests are shown above the brackets. IRT=immune reconstitution therapy, miRNA=microRNA, MS=multiple sclerosis, n=number, PPMS=primary progressive multiple sclerosis, RRMS=relapsing-remitting multiple sclerosis.

rs1005039 and rs817478 directly implicate an altered processing of hsa-mir-199a-1 and hsa-mir-4423, respectively. The rare MS risk allele of SNP rs1005039 leads to

an A in the second of two adjacent CNNC motifs in the transcript sequence (Figure 4a). In general, a stronger decrease in relative luminescence was observed when

MIR SNP	Allele frequency	RA	MicroRNA assay <sup>a</sup>	0 RA, mean $\pm$ SD (mv <sup>b</sup> )	1 RA, mean $\pm$ SD (mv <sup>b</sup> )	2 RA, mean $\pm$ SD (mv <sup>b</sup> )	P-value <sup>c</sup>
rs41292017	G: 98.9%; A: 1.1%	A	hsa-miR-26a-2-5p[hsa-miR-26a-5p] 13_34	2.01 $\pm$ 0.44	2.07 $\pm$ 0.64	—	0.5902
			hsa-miR-26a-2-5p[isomiR] 13_33	0.65 $\pm$ 0.53	0.78 $\pm$ 0.85	—	0.5059
			hsa-miR-26a-2-3p[hsa-miR-26a-2-3p] 51_72	13.51 $\pm$ 1.15 (10)	13.69 $\pm$ 0.83	—	0.4343
rs1005039	C: 97.8%; T: 2.2%	T	hsa-miR-199a-1-5p[hsa-miR-199a-5p] 5_27	21.07 $\pm$ 5.77 (59)	17.98 $\pm$ 5.64	—	0.7023
			hsa-miR-199a-1-5p[isomiR] 5_26	11.63 $\pm$ 2.30 (2)	10.48 $\pm$ 2.51	—	0.3546
			hsa-miR-199a-1-3p[hsa-miR-199a-3p] 46_67	7.78 $\pm$ 1.82	6.71 $\pm$ 2.23	—	0.1580
			hsa-miR-199a-1-3p[isomiR] 46_66	8.00 $\pm$ 1.46	7.07 $\pm$ 1.79	—	0.2090
rs78351440	C: 94.5%; A: 5.5%	C	hsa-miR-4304-5p[hsa-miR-4304] 10_26	—	10.54 $\pm$ 0.69	10.68 $\pm$ 0.88	0.7112
			hsa-miR-4304-3p[isomiR] 35_52	—	10.93 $\pm$ 1.09	10.98 $\pm$ 0.97	0.7410
rs817478	A: 84.6%; C: 15.4%	A	hsa-miR-4423-5p[hsa-miR-4423-5p] 12_33	17.56 $\pm$ 4.51 (1)	20.46 $\pm$ 5.02 (8)	21.99 $\pm$ 4.41 (46)	<b>0.0015</b>
			hsa-miR-4423-5p[isomiR] 12_32	18.55 $\pm$ 5.01 (2)	19.02 $\pm$ 4.15 (8)	21.13 $\pm$ 3.87 (41)	<b>0.0030</b>
			hsa-miR-4423-3p[hsa-miR-4423-3p] 48_68	12.75 $\pm$ 0.92	13.16 $\pm$ 0.90	13.52 $\pm$ 1.04	0.0934
			hsa-miR-4423-3p[isomiR] 48_67	12.68 $\pm$ 0.68	12.79 $\pm$ 0.95	13.16 $\pm$ 0.91	0.1230
rs7926599	G: 98.9%; C: 1.1%	G	hsa-miR-4492-3p[isomiR] 58_74	—	8.30 $\pm$ 0.43	6.59 $\pm$ 1.18	0.5171

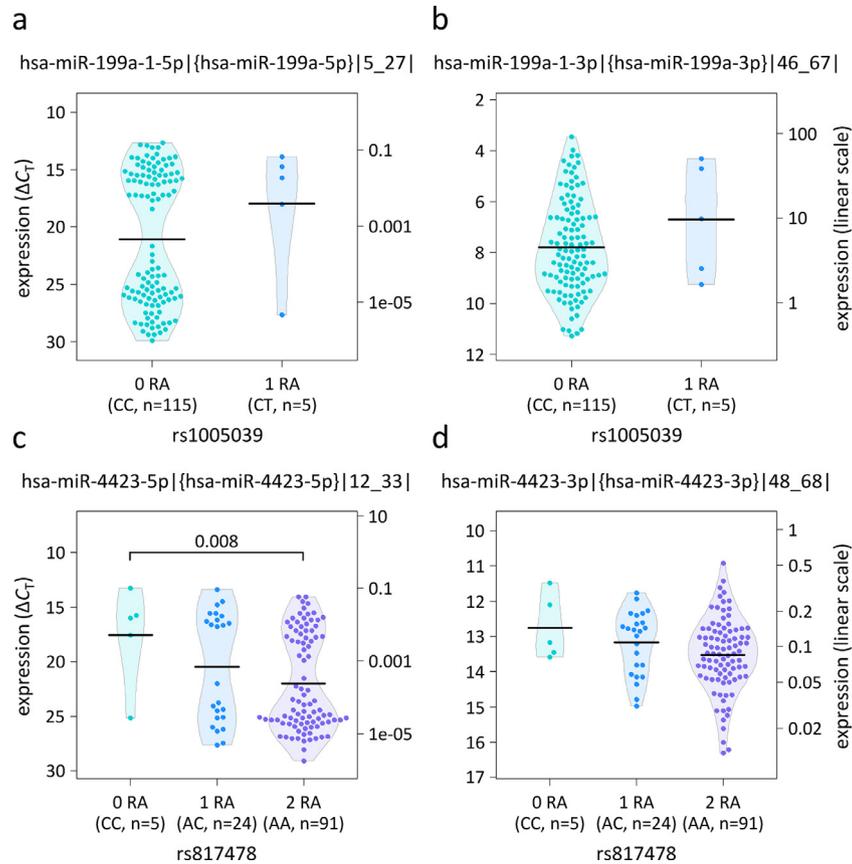
**Table 4: Results of the analysis to detect *cis*-miR-eQTL effects in B cells.**

The expression of mature miRNA molecules in B cells from the peripheral blood was measured by stem-loop qPCR assays. The *cis*-miR-eQTL analysis included only miRNAs that could be detected in at least half of the samples (n=120) and only MS-associated SNPs in miRNA-coding regions (MIR SNPs) for which both alleles were present in the data. For each MIR SNP, the allele frequencies in the study population (n=91 subjects) as well as the respective MS risk allele are given in the table. Average  $\Delta C_T$  values are reported dependent on the MIR SNP genotype as specified by the number of risk alleles carried. Lower  $\Delta C_T$  values correspond to higher expression levels. For 4 of the 5 MIR SNPs, there were no homozygotes for the minor allele (—). Significance values indicating *cis*-miR-eQTL effects ( $P < 0.05$ ) are marked in bold. — = not available, eQTL=expression quantitative trait locus, MIR/miRNA= microRNA, MS= multiple sclerosis, mv=missing values, qPCR=quantitative polymerase chain reaction, RA=risk allele, SD=standard deviation, SNP=single-nucleotide polymorphism.

<sup>a</sup> the miRNAs are specified following the nomenclature by Cloonan et al.<sup>36</sup>

<sup>b</sup> the number of samples in which the miRNA could not be detected within 45 qPCR cycles and for which  $C_T$  values were thus imputed.

<sup>c</sup> Wald  $\chi^2$  test P-values for linear mixed-effects models fitted to the data in linear scale and adjusted for age, sex and group assignment.



**Figure 3.** Levels of hsa-miR-199a-5p and -3p as well as hsa-miR-4423-5p and -3p per genotype.

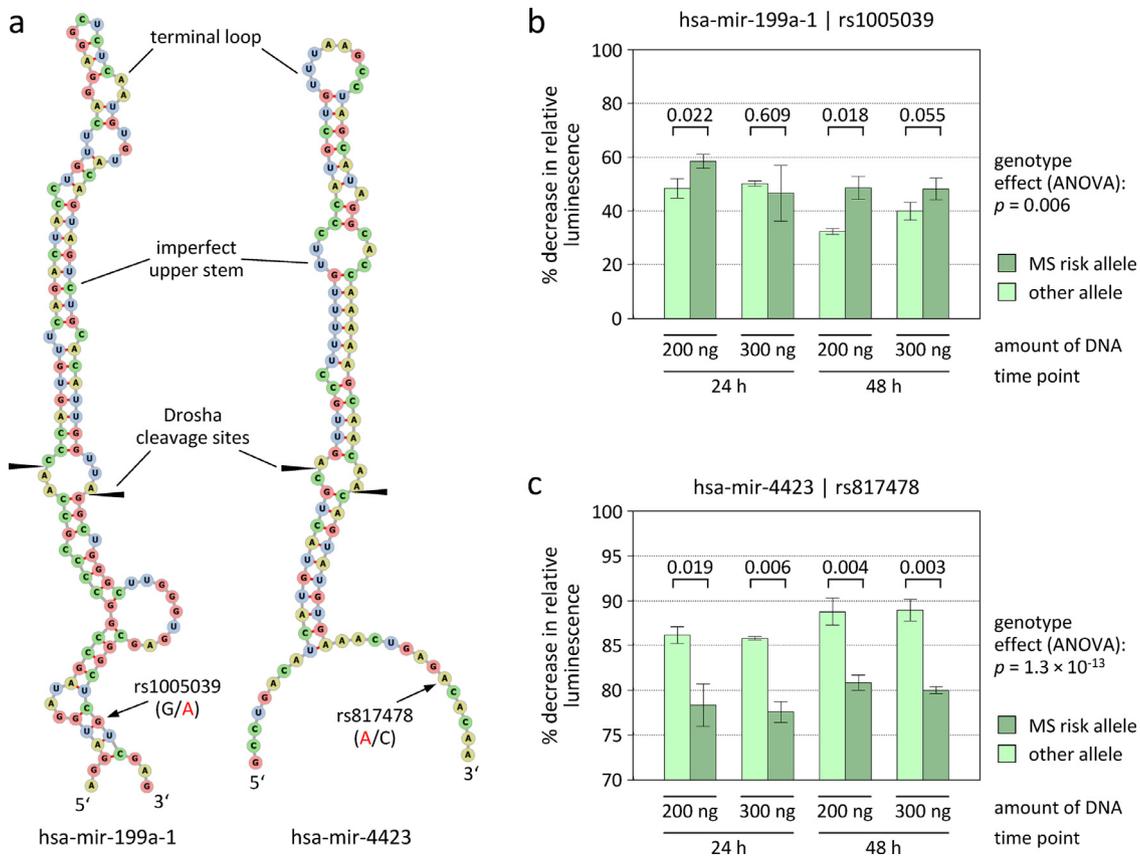
The B-cell expression levels of 4 miRBase-annotated mature miRNAs are visualised. (a and b) A non-significant tendency towards higher expression of hsa-miR-199a-5p and -3p was noted in carriers of the MS risk allele (RA) of SNP rs1005039. (c and d) The RA of SNP rs817478 was associated with a lower expression of hsa-miR-4423-5p in the *cis*-miR-eQTL analysis (Table 4). Accordingly, the respective SNPs were suspected to affect the miRNA stem-loop processing by Drosha. The number of samples from subjects carrying 0, 1 or 2 RA is indicated below each beeswarm/violin plot. The left and right y-axes refer to the raw qPCR data ( $\Delta C_T$  values displayed in an inverted manner) and the converted data ( $2^{-\Delta C_T} \times 1,000$ ), respectively. The means per genotype group are shown as black lines. The Tukey test P-value reaching the significance level of  $\alpha=0.05$  is indicated above the bracket. eQTL=expression quantitative trait locus, miRNA=microRNA, MS=multiple sclerosis, n=number, qPCR=quantitative polymerase chain reaction, SNP=single-nucleotide polymorphism.

HeLa cells were transfected with the plasmid carrying the risk allele than when HeLa cells were transfected with the plasmid carrying the other allele (Figure 4b). Although this difference was not always statistically significant at individual experimental conditions in *t*-tests, the overall genotype effect was significant when analysing the whole dataset in a 3-way ANOVA ( $P=0.006$ ). This points to a more efficient processing of the hsa-mir-199a-1 stem-loop in the presence of the risk allele of SNP rs1005039.

The SNP rs817478 denotes an A/C polymorphism. The C allele creates a CNNC motif downstream of hsa-mir-4423, whereas the motif is missing when the more common MS risk allele A is present. In the luciferase reporter assay, a strong decrease in relative

luminescence (>75%) was generally seen after transfection of HeLa cells with GLuc-hsa-mir-4423 constructs. However, a significantly lower decrease was consistently observed when the construct carrying the risk allele was used (*t*-test  $P \leq 0.019$  and ANOVA  $P=1.3 \times 10^{-13}$ ). This clearly indicates that the MS risk allele of SNP rs817478 is associated with a much lower efficiency of hsa-mir-4423 processing by Drosha (Figure 4c).

We also assessed whether the SNPs rs41292017, rs78351440 and rs7926599 affect the cleavage of hsa-mir-26a-2, hsa-mir-4304 and hsa-mir-4492, respectively. However, in line with the results of the *cis*-miR-eQTL analysis (Table 4), there was no evidence of a genotype-dependent processing of these miRNA stem-loops in the reporter assays (3-way ANOVA  $P > 0.125$ ).



**Figure 4.** Allele-specific processing of microRNA precursor sequences in the reporter assay.

(a) The predicted secondary structures of the miRNA stem-loops were visualised using the web-based tool forna.<sup>58</sup> Both SNPs are located in a CNNC motif in the 3' flanking region. The MS risk allele (RA) is marked in red. The Drosha cleavage sites are indicated according to the miRBase annotation.<sup>36</sup> (b) and (c) The processing of hsa-mir-199a-1 and hsa-mir-4423 was analysed using a luciferase-based assay. For this purpose, the miRNA precursor sequences were cloned into the 3' UTR of the GLuc gene within a reporter vector. GLuc/SEAP ratios were then measured after transient transfection of HeLa cells with plasmids carrying either the MS risk allele or the alternative allele. The decrease in relative luminescence was calculated by subtracting from 1 the quotient of the GLuc/SEAP ratio and the respective ratio that was obtained for the negative control. Higher bars thus indicate higher processing efficiencies. The bars and error bars show the means and standard deviations of 3 biological replicates each. Welch *t*-test P-values are given above the bars. The MS RA of the SNP rs1005039 conferred a higher hsa-mir-199a-1 cleavage efficiency (b), whereas the RA of SNP rs817478 was associated with a significantly diminished processing of hsa-mir-4423, independently of the amount of plasmid DNA used for transfection and the readout time point (c). ANOVA=analysis of variance (3-way additive model), GLuc=*Gaussia* luciferase, miRNA=microRNA, MS=multiple sclerosis, SEAP=secreted alkaline phosphatase, SNP=single-nucleotide polymorphism, UTR=untranslated region.

### Target genes of hsa-mir-199a-1 and hsa-mir-4423

The *cis*-miR-eQTL analyses and the luciferase reporter assays suggested that MS-associated SNPs modulate the processing of hsa-mir-199a-1 and hsa-mir-4423. To identify their target genes, we transfected HeLa cells either with the respective precursor miRNA expression plasmids, which did not carry the MS risk allele but the other allele, or with the scrambled control plasmid. The qPCR analysis revealed that the endogenous expression of the mature miRNAs in HeLa cells was low or absent. In contrast, a massive overexpression of the miRNAs was detected 24 h and 48 h after transfection with the precursor miRNA plasmids (Table S3).

The subsequent microarray analysis disclosed the expression levels of 135,750 different TCs. The transcriptome data were of high quality and showed a similar distribution across all 18 samples (Figure S3). Compared to the negative controls, 152 TCs were expressed at significantly lower levels when hsa-mir-199a-1 was overexpressed ( $P < 0.05$  and fold-change  $< -1.5$ ). Of those, 60 transcripts were found to encode proteins and to contain at least one predicted binding site for the respective 5p or 3p mature miRNAs in their 3' UTRs (Table 5). For HeLa cells that were transfected with the hsa-mir-4423 plasmid, we filtered 62 TCs with lower expression ( $P < 0.05$  and fold-change  $< -1.5$ ). However, of those,

Primary microRNA	N	Prioritised target genes	Gene functions
hsa-mir-199a-1	60	ACTN4, AMZ2, ARHGAP31, ARPC1A, BMP1, CARD19, CD9, CDK4, CIZ1, CSNK1G2, CSRNP1, DYRK1B, EEF2, EFNA1, EPST11, FAIM, FEN1, FXR2, HEATR3, HTRA2, IER3, IRAK2, MAP3K12, MIF, MLX, MRPS36, MTG2, NACC1, NDUFA11, NT5C, OGFOD3, OR8B3, PACS1, PEA15, PFKP, PNPLA6, PON2, PPIB, PRMT1, PROCR, RAB8A, RAD9A, RAP2B, RPS29, SIRT6, SPATA2, TAC1, TCEAL3, TMOD2, TNFRSF12A, TPRA1, TRAF1, TUBB6, UBE2D1, VIM, ZDHHC3, ZFYVE27, ZNF511, ZNF526, ZNF579	Reactome pathways: Disease, Immune system, Infectious disease, Metabolism, Signal transduction Gene Ontology categories: Cellular protein modification process, RNA binding
hsa-mir-4423	1	TMEM47	UniProt: Cell junction organisation

**Table 5: Prioritised target genes of hsa-mir-199a-1 and hsa-mir-4423.**

We obtained transcriptome profiles of HeLa cells in which the respective miRNA precursor transcripts were overexpressed. Listed are the potential direct target genes identified by this screening that encode proteins and for which a hybridisation of their 3' UTR sequences with the mature miRNA sequences could be computationally verified. The online tools Reactome<sup>79</sup> and Enrichr<sup>80</sup> were used to analyse the involvement of the target genes of hsa-mir-199a-1 in biological processes. See Table S4 and Table S6 for further information. miRNA=miRNA, N=number, UTR=untranslated region.

*TMEM47* was the only protein-coding transcript with sequence complementarity in the 3' UTR to the corresponding mature miRNAs. The mRNA expression fold-changes and P-values as well as the estimated minimum hybridisation energies are provided for all prioritised target genes in Table S4. Information on their expression in cell types of the blood and brain is given in Table S5. The transcriptome data are available in the GEO database (accession number GSE185859).

A subset of 26 of the 60 target genes that we identified for hsa-mir-199a-1 were also predicted as putative targets in the miRWalk database,<sup>77</sup> and *NACC1* and *TMOD2* are already listed as experimentally verified targets in the miRTarBase database.<sup>78</sup> The target genes of hsa-mir-199a-1 were found to participate in diverse biological functions, including immune system signalling, metabolism and RNA binding. *TMEM47*, the prioritised target gene of hsa-mir-4423, is a transmembrane protein that regulates the organisation of cellular junctions. Interactions of hsa-miR-4423-5p and -3p with the 3' UTR of *TMEM47* were also predicted by miRWalk<sup>77</sup> (Tables 5 and S6).

## Discussion

Until now, about 200 genetic loci have been linked to the inherited risk of developing MS,<sup>26</sup> but our functional understanding of the underlying molecular mechanisms is still scarce. Here, we hypothesised that pri-miRNA processing sites contain causal variants for this disease. We combined expression analyses of B cells, experiments in cell culture and *in silico* evaluations to shed light on the maturation and function of largely underexplored and primate-specific miRNAs. To our knowledge, at least the prioritised miRNAs with 4-digit numbers have never been explicitly measured in B cells from MS patients before. We found that the Drosha-

mediated cleavage of hsa-mir-199a-1 and, in particular, hsa-mir-4423 is influenced by proximal MS risk variants. We have also determined potential direct target genes of these miRNAs.

We first examined the association between the genotype of MS-associated SNPs and the levels of mature miRNAs and isomiRs in B cells. This was done in a representative study population, with the exception that we did not include patients with SPMS. Otherwise, the patient cohort reflected typical characteristics of MS patients in terms of age, sex and disease status, resembling those of patients included in European MS registries.<sup>88</sup> With flow cytometry, we confirmed a strong enrichment of B cells in the samples, while the percentages of the 8 B-cell subpopulations were within the expected ranges: The distribution of B-cell subsets was quite similar when comparing healthy subjects with the group of RRMS patients before IRT. Patients receiving natalizumab infusions had a lower proportion of naive B cells but a higher proportion of memory B cells in the blood, as previously described.<sup>89,90</sup> In contrast, patients who were treated with oral cladribine or intravenous alemtuzumab showed a relative increase in the frequency of transitional and naive B cells and a decrease in the frequency of memory B cells. This is in line with earlier studies that have shown that B-cell repopulation following infusion of alemtuzumab or (subcutaneous) administration of cladribine results in such shifts in B-cell subpopulations.<sup>15,16,91,92</sup> However, the effects of these IRTs for MS on miRNA expression have not been investigated so far.

Compared to the healthy controls, we observed a significant differential expression of hsa-miR-26a-5p and -3p in RRMS patients treated with natalizumab and alemtuzumab, respectively. In previous studies, altered levels of hsa-miR-26a-5p were detected in serum,<sup>93,94</sup> whole blood,<sup>95</sup> PBMC<sup>96–98</sup> and thrombocytes<sup>99</sup> of

patients with MS. However, inconsistent results have been described, most likely due to different study designs. For instance, in response to interferon- $\beta$  therapy, higher levels of this miRNA were reported in thrombocytes,<sup>99</sup> but lower levels were reported in serum exosomes.<sup>94</sup> Our study differs from the existent studies in that we explicitly analysed B cells, that we also measured hsa-miR-26a-3p levels and that we included patients who received IRTs. The pri-miRNA hsa-mir-26a-2 is encoded on chromosome 12 and belongs to the mir-26 family, which is conserved in vertebrates.<sup>36</sup> Of note, hsa-miR-26a-5p is also produced from the hsa-mir-26a-1 locus on chromosome 3, which weakened the validity of the *cis*-miR-eQTL analysis. There is accumulating evidence that miR-26a miRNAs play pivotal roles in cellular differentiation, cell growth and apoptosis and that they are involved in the development of various human diseases.<sup>100</sup> In experimental autoimmune encephalomyelitis (EAE), an animal model of MS, silencing of miR-26a-5p was found to result in increased expression of T<sub>H</sub>17-related cytokines and increased disease severity, while overexpression of the precursor miRNA was found to result in reduced expression of T<sub>H</sub>17-related cytokines and a milder form of EAE.<sup>95</sup> A direct target of miR-26a-5p is IL6 mRNA.<sup>95</sup>

We measured higher levels of hsa-miR-199a-5p and -3p in B cells from alemtuzumab-treated patients than in those from healthy controls. The literature on the expression of these miRNAs in MS patients is heterogeneous, with previous studies performed with sera,<sup>101</sup> PBMC,<sup>102</sup> CD4<sup>+</sup> T cells,<sup>103</sup> cerebrospinal fluid samples<sup>104</sup> and brain lesion tissues.<sup>105</sup> Remarkably, the levels of hsa-miR-199a-5p were described to be ~3 times higher in CD4<sup>+</sup> T cells of RRMS patients in remission compared to healthy subjects<sup>103</sup> and in MS lesions compared to normal brain white matter.<sup>105</sup> The mir-199 family is evolutionarily highly conserved, and it should be noted that the -5p and -3p mature miRNA molecules are produced from two loci (hsa-mir-199a-1 on chromosome 19 and hsa-mir-199a-2 on chromosome 1).<sup>36</sup> Functional studies have demonstrated that especially the -3p mature form may act as either a tumour suppressor or an oncogene, depending on the type of cancer.<sup>106</sup> The miR-199a miRNAs were shown to regulate tumour cell proliferation, invasion, apoptosis and glucose metabolism, to protect against hypoxia-induced damage in cardiomyocytes, to modulate stem cell differentiation and fetal development and to affect adipocyte differentiation and fatty acid composition.<sup>106–108</sup>

In our study, only 5 B-cell samples were heterozygous for the MS-associated allele T of SNP rs1005039 that is located in one of two CNNC motifs flanking hsa-mir-199a-1. The low MAF possibly accounts for the fact that the suspected *cis*-miR-eQTL effect could not be detected with statistical significance. However, the results of the luciferase reporter assays indicated a higher pri-miRNA processing efficiency in HeLa cells

when the MS risk allele was present. Likewise, Roden *et al.* have demonstrated that an exchange of the NN bases within CNNC can have a mild effect on the stem-loop processing.<sup>47</sup> Moreover, a previous study reported that the C allele of SNP rs3760781, which has a distance of ~19 kb to the pri-miRNA locus, is associated with significantly higher levels of hsa-miR-199a-3p and -5p in plasma.<sup>109</sup> The worldwide frequency of the haplotype TCC of the MIR SNP (rs1005039), the SNP from<sup>109</sup> (rs3760781) and the MS lead SNP (rs12609500) is 1.3% according to the 1000 Genomes data.<sup>51,52</sup> In such a context, a small sample size, as in our study, implies high uncertainty in the inferences that can be drawn.<sup>110</sup> Hence, to further substantiate the presumed causal relationship between genetic risk, miRNA expression and MS, a larger study population and/or a family-based approach is needed. A number of target genes of miR-199a miRNAs have been already experimentally determined in previous studies, as recorded in the miRTarBase database.<sup>78</sup> In the present study, 60 protein-coding target genes were prioritised for hsa-mir-199a-1, some of which are novel. For instance, the immune-related genes *IRAK2*, *MIF* and *VIM* were not contained as known or predicted targets in the databases miRTarBase<sup>78</sup> and miRWalk,<sup>77</sup> respectively. These and other target gene candidates that were identified by our transcriptome screening await further confirmation in future studies.

The MS risk allele A of SNP rs817478 was associated with significantly lower levels of hsa-miR-4423-5p in B cells in our *cis*-miR-eQTL analysis. The same trend was seen for hsa-miR-4423-3p. In the reporter assay experiment, an allele-dependent cleavage rate of hsa-mir-4423 was clearly confirmed across all experimental conditions. These data thus suggest a causal role of miR-4423 miRNAs in the genetic predisposition to MS. The association of the rs817478 genotype with hsa-miR-4423-5p expression was previously identified as one of 57 *cis*-miR-eQTLs in the European subset of the Geuvadis RNA-sequencing data on EBV-transformed lymphoblastoid cell lines.<sup>33</sup> The eQTL is also contained in the ncRNA-eQTL database, which is solely based on cancer-related studies.<sup>111</sup> The genetic effect is explained by the fact that there is a 3' flanking CNNC motif in the presence of the minor allele C but not in the presence of the MS risk allele A. This sequence feature is well known to enhance the Drosha-mediated processing and is found downstream of one third of all human conserved miRNA stem-loops.<sup>44,47</sup> The CNNC motif is bound by SRSF3, which is a major auxiliary factor for pri-miRNA processing,<sup>112</sup> and both hsa-mir-199a-1 and hsa-mir-4423 were found to be productively processed by Drosha in the presence of SRSF3.<sup>113</sup> The pri-miRNA hsa-mir-4423 has recently evolved: It is currently only annotated in the genomes of humans and chimpanzees.<sup>36</sup> The mature miRNA from the 3' arm has been first identified in malignant B cells.<sup>114</sup> Perdomo *et al.* have shown that

hsa-mir-4423 is a regulator of airway epithelial cell differentiation and a repressor of lung carcinogenesis.<sup>115</sup> Otherwise, its function is still largely unclear. The miR-TarBase database currently contains no miRNA-target interaction with strong evidence (*i.e.*, with verification by qPCR, western blot or reporter assay) for hsa-mir-4423.<sup>78</sup> Our analysis revealed only *TMEM47* as potential direct target gene of this miRNA. This gene was also found to be significantly lower expressed in SW900 cells overexpressing hsa-mir-4423 in the data from Perdomo et al. (GEO accession GSE48796).<sup>115</sup> *TMEM47* is an adherens junction protein involved in regulating epithelial cell junction organisation.<sup>116</sup> It is also upregulated during astrocyte maturation,<sup>117</sup> but it remains to be clarified to which degree it plays a role in MS. Further investigations on the cell type-specific functions of this gene are warranted.

Several limitations of this study should be considered in the data interpretation. First, the patients in our study were heterogeneous with respect to the treatment they received. We accounted for this by including the group assignment in the *cis*-miR-eQTL analysis, while also adjusting for age and sex, but the possibility of unmeasured confounding cannot be ruled out. It also remains unclear whether other IRTs, such as ocrelizumab, affect the expression of the miRNAs. Second, the measurement of mature miRNA levels was conducted in B cells, and we restricted our analysis to miRNAs that could be detected in at least half of the samples. However, miRNAs from hsa-mir-3661 and hsa-mir-4252 could not be assayed properly, and other miRNAs were barely detectable, including hsa-miR-548ac-3p, for which we have previously shown a *cis*-miR-eQTL effect and which is preferentially expressed in CD8<sup>+</sup> T cells and natural killer cells.<sup>48,118</sup> We therefore presumably missed eQTLs for miRNAs that are expressed predominantly in other cells as well as eQTLs that are cell type-specific, a phenomenon that has been studied more thoroughly for protein-coding genes.<sup>119,120</sup> Deeper insights on the impact of the genetic risk variants might be achieved by examining other peripheral immune cells, CNS-resident cells and/or specific B-cell subsets, *e.g.*, memory B cells. Third, our study was limited to selected SNPs and miRNAs. We did not include rare variants with MAF < 1% because the sample size would have been insufficient to study *cis*-miR-eQTL associations. We also excluded miRNAs that are produced independently of Drosha (*e.g.*, mirtrons<sup>39</sup>). Moreover, it is difficult to define possible isomiRs and to measure their expression accurately. Here, we employed stem-loop RT-qPCR assays. These assays allow to discriminate between closely related mature miRNAs,<sup>121</sup> but different 3' isomiR species may be cross-detected.<sup>122</sup> The exclusive measurement of miRNA isoforms that differ by only 1 nt at their termini is possible by dumbbell PCR with TaqMan probes,<sup>123</sup> but there are currently no commercial assays for this. Fourth, the efficiency of pri-

miRNA processing was evaluated using reporter assays in which transiently expressed luciferase mRNA is cleaved by endogenous Microprocessor in HeLa cells.<sup>69,70</sup> As an alternative, recombinant Drosha and DGCR8 proteins can be used in kinetic assays to more precisely determine the allele-specific rates at which the cleavage occurs.<sup>124</sup> It is important to note, however, that such assays do not reflect all regulatory events taking place *in vivo*. In fact, various mechanisms coordinate the biogenesis and turnover of mature miRNAs (*e.g.*, Dicer activity and miRNA decay<sup>39</sup>). Further research is needed to elucidate how the interplay of genetic factors with environmental and lifestyle factors (*e.g.*, EBV infection and vitamin D deficiency) may alter molecular networks related to miRNA maturation and function.<sup>125,126</sup> Fifth, a transcriptome-wide analysis was conducted to identify the most likely direct target genes of hsa-mir-199a-1 and hsa-mir-4423. However, we might have missed actual target genes, *e.g.*, those that are not endogenously expressed in HeLa cells and those that are inhibited by translational repression rather than mRNA destabilisation. In future studies, improved experimental strategies for miRNA target identification may be applied to verify effects at the protein level and to better resolve the functional differences between miRNA family members and isoforms.<sup>127</sup>

In conclusion, we here focused on miRNAs that are encoded in MS-associated genetic regions. We observed a significant differential expression of mature miRNAs and isomiRs from hsa-mir-26a-2 and hsa-mir-199a-1 in B cells when comparing subgroups of RRMS patients and healthy controls. Based on *cis*-miR-eQTL analyses and reporter assays, we found that the MS risk allele T of SNP rs1005039 is related to a somewhat more efficient Drosha-dependent processing of hsa-mir-199a-1, whereas the MS risk allele A of SNP rs817478 confers a much lower hsa-mir-4423 processing efficiency. These findings complement the existing body of literature on the dysregulation of miRNAs in MS and on sequence determinants of pri-miRNA processing. Our and other data indicate that diverse biological processes are regulated by hsa-mir-199a-1, while the role of hsa-mir-4423 remains poorly defined. For a deeper understanding of the pathomechanisms of MS, further functional studies are needed to identify the disease-causing genetic variants and to ascertain their effects on transcriptional activity and RNA processing.

#### Contributors

MH, BF and UKZ conceptualised the study. MH, NB, EP and UKZ secured the research funding. MS, AW, SM and AD coordinated the blood sampling and collected clinical information. NB, EP and BF processed the blood samples. NB executed the majority of the experiments. DK performed the microarray analysis. PL provided valuable suggestions for the reporter assays.

NB and MH analysed and interpreted the data and prepared the figures and tables. UKZ supervised the research and provided important intellectual content. MH drafted the original manuscript, and all authors reviewed, edited and prepared the final version. MH and NB had full access to all the data and take responsibility for the integrity and accuracy of the data analysis. EP, BF and DK have verified the underlying data. All authors have read and approved the final version of the manuscript.

### Data sharing statement

The data supporting the findings of this study are available in the article and/or supplementary materials. The transcriptome data from the target gene analysis are publicly available from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE185859>). Readers are welcome to contact the corresponding author by e-mail to request access to the raw de-identified data used in this work.

### Declaration of interests

MH received speaking fees and travel funds from Bayer HealthCare, Biogen, Merck, Novartis and Teva. AW received speaking fees and travel funds from Biogen, GlaxoSmithKline, Merck Serono, Novartis and Sanofi Genzyme. NB received travel funds from Novartis. UKZ received research support as well as speaking fees and travel funds from Alexion, Almirall, Bayer HealthCare, Biogen, Bristol Myers Squibb, Janssen, Merck Serono, Novartis, Roche, Sanofi Genzyme, Teva as well as EU, BMBF, BMWi and DFG. BF, EP, MS, SM, AD, DK and PL declare that they have no competing interests.

### Acknowledgments

We thank Antje Bombor and Ina Schröder for coordinating the clinical visits and blood draws, respectively. We are grateful to Nele Retzlaff and Deborah Sonnenberg for reviewing the medical records. We wish to thank Hans-Jürgen Thiesen and Robert Jaster for providing access to technical equipment. We thank Ildikó Tóth for laboratory support. We would like to thank Robby Engelmann as well as Wendy Bergmann, Michael Müller and Johann Volzke from the Core Facility for Cell Sorting and Cell Analysis at the Rostock University Medical Center for their kind assistance in the flow cytometric analyses. We thank Felix Steinbeck for his input on computational tools for analysing genetic data.

This study was funded by the Rostock University Medical Center (FORUN program, grant: 889002), Sanofi Genzyme (grant: GZ-2016-11560) and Merck Serono GmbH (Darmstadt, Germany, an affiliate of Merck KGaA, CrossRef Funder ID: 10.13039/

100009945, grant: 4501860307). NB was supported by the Stiftung der Deutschen Wirtschaft (sdw) and the FAZIT foundation. EP was supported by the Landesgraduiertenförderung Mecklenburg-Vorpommern.

### Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2022.104052.

### References

- Filippi M, Bar-Or A, Piehl F, et al. Multiple sclerosis. *Nat Rev Dis Prim*. 2018;4(1):43.
- Filippi M, Brück W, Chard D, et al. Association between pathological and MRI findings in multiple sclerosis. *Lancet Neurol*. 2019;18(2):198–210.
- Comi G, Bar-Or A, Lassmann H, et al. Role of B cells in multiple sclerosis and related disorders. *Ann Neurol*. 2021;89(1):13–23.
- Zettl UK, Stüve O, Patejdl R. Immune-mediated CNS diseases: a review on nosological classification and clinical features. *Autoimmun Rev*. 2012;11(3):167–173.
- Rommer PS, Eichstädt K, Ellenberger D, et al. Symptomatology and symptomatic treatment in multiple sclerosis: results from a nationwide MS registry. *Mult Scler J*. 2019;25(12):1641–1652.
- Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology*. 2014;83(3):278–286.
- Frahm N, Ellenberger D, Fneish F, et al. Characteristics of secondary progressive multiple sclerosis: disease activity and provision of care in Germany-A registry-based/multicentric cohort study. *Mult Scler Relat Disord*. 2021;56: 103281.
- Frahm N, Peters M, Bätzing J, et al. Prevalence of pediatric multiple sclerosis in Germany: a nationwide population-based analysis. *Eur J Neurol*. 2021;28(9):3173–3176.
- Ribbons KA, McElduff P, Boz C, et al. Male sex is independently associated with faster disability accumulation in relapse-onset ms but not in primary progressive MS. *PLoS One*. 2015;10(6): e0122686.
- Khoy K, Mariotte D, Defer G, Petit G, Toutirais O, Le Mauff B. Natalizumab in multiple sclerosis treatment: from biological effects to immune monitoring. *Front Immunol*. 2020;11: 549842.
- Rommer P, Zettl UK. Treatment options in multiple sclerosis and neuromyelitis optica spectrum disorders. *Curr Pharm Des*. 2022;28(6):428–436.
- Lünemann JD, Ruck T, Muraro PA, Bar-Or A, Wiendl H. Immune reconstitution therapies: concepts for durable remission in multiple sclerosis. *Nat Rev Neurol*. 2020;16(1):56–62.
- Sorensen PS, Sellebjerg F. Pulsed immune reconstitution therapy in multiple sclerosis. *Ther Adv Neurol Disord*. 2019;12: 1756286419836913.
- Baker D, Herrod SS, Alvarez-Gonzalez C, Zaleski L, Albor C, Schmierer K. Both cladribine and alemtuzumab may effect MS via B-cell depletion. *Neurol Neuroimmunol Neuroinflamm*. 2017;4(4): e360.
- Baker D, Herrod SS, Alvarez-Gonzalez C, Giovannoni G, Schmierer K. Interpreting lymphocyte reconstitution data from the pivotal phase 3 trials of alemtuzumab. *JAMA Neurol*. 2017;74(8):961–969.
- Baker D, Pryce G, Herrod SS, Schmierer K. Potential mechanisms of action related to the efficacy and safety of cladribine. *Mult Scler Relat Disord*. 2019;30:176–186.
- Hauser SL, Cree BAC. Treatment of multiple sclerosis: a review. *Am J Med*. 2020;133(12):1380–1390.
- Rommer PS, Milo R, Han MH, et al. Immunological aspects of approved MS therapeutics. *Front Immunol*. 2019;10:1564.
- Cencioni MT, Mattosio M, Magliozzi R, Bar-Or A, Muraro PA. B cells in multiple sclerosis - from targeted depletion to immune reconstitution therapies. *Nat Rev Neurol*. 2021;17(7):399–414.
- von Büdingen HC, Palanichamy A, Lehmann-Horn K, Michel BA, Zamvil SS. Update on the autoimmune pathology of multiple sclerosis: B-cells as disease-drivers and therapeutic targets. *Eur Neurol*. 2015;73(3-4):238–246.

- 21 Li R, Patterson KR, Bar-Or A. Reassessing B cell contributions in multiple sclerosis. *Nat Immunol*. 2018;19(7):696–707.
- 22 Olsson T, Barcellos LF, Alfredsson L. Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis. *Nat Rev Neurol*. 2017;13(1):25–36.
- 23 Hecker M, Bühring J, Fitzner B, Rommer PS, Zettl UK. Genetic, environmental and lifestyle determinants of accelerated telomere attrition as contributors to risk and severity of multiple sclerosis. *Biomolecules*. 2021;11(10):1510.
- 24 Moutsianas L, Jostins L, Beecham AH, et al. Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nat Genet*. 2015;47(10):1107–1113.
- 25 Cotsapas C, Mitrovic M. Genome-wide association studies of multiple sclerosis. *Clin Transl Immunol*. 2018;7(6):e1018.
- 26 International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science*. 2019;365(6460):eaav7188.
- 27 Guo MH, Sama P, LaBarre BA, et al. Dissection of multiple sclerosis genetics identifies B and CD4+ T cells as driver cell subsets. *bioRxiv*. 2021. <https://doi.org/10.1101/2021.05.24.445445>.
- 28 Cano-Gamez E, Trynka G. From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. *Front Genet*. 2020;11:424.
- 29 Vandiedonck C. Genetic association of molecular traits: a help to identify causative variants in complex diseases. *Clin Genet*. 2018;93(3):520–532.
- 30 GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369(6509):1318–1330.
- 31 Kerimov N, Hayhurst JD, Peikova K, et al. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat Genet*. 2021;53(9):1290–1299.
- 32 Ota M, Nagafuchi Y, Hatano H, et al. Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell*. 2021;184(11):3006–3021.
- 33 Lappalainen T, Sammeth M, Friedländer MR, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*. 2013;501(7468):506–511. Sep 26.
- 34 Huan T, Rong J, Liu C, et al. Genome-wide identification of microRNA expression quantitative trait loci. *Nat Commun*. 2015;6:6601.
- 35 Bartel DP. Metazoan microRNAs. *Cell*. 2018;173(1):20–51.
- 36 Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res*. 2019;47(D1):D155–D162.
- 37 Boivin V, Deschamps-Francoeur G, Scott MS. Protein coding genes as hosts for noncoding RNA expression. *Semin Cell Dev Biol*. 2018;75:3–12.
- 38 Treiber T, Treiber N, Meister G. Regulation of microRNA biogenesis and its crosstalk with other cellular pathways. *Nat Rev Mol Cell Biol*. 2019;20(1):5–20.
- 39 Ha M, Kim VN. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol*. 2014;15(8):509–524.
- 40 Feng Y, Zhang X, Song Q, Li T, Zeng Y. Drosha processing controls the specificity and efficiency of global microRNA expression. *Biochim Biophys Acta*. 2011;1809(11–12):700–707.
- 41 Conrad T, Marsico A, Gehre M, Orom UA. Microprocessor activity controls differential miRNA biogenesis *In Vivo*. *Cell Rep*. 2014;9(2):542–554.
- 42 Bofill-De Ros X, Yang A, Gu S. IsomiRs: expanding the miRNA repression toolbox beyond the seed. *Biochim Biophys Acta Gene Regul Mech*. 2020;1863(4):194373.
- 43 Creugny A, Fender A, Pfeffer S. Regulation of primary microRNA processing. *FEBS Lett*. 2018;592(12):1980–1996.
- 44 Auyeung VC, Ulitsky I, McGeary SE, Bartel DP. Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell*. 2013;152(4):844–858.
- 45 Mori M, Triboulet R, Mohseni M, et al. Hippo signaling regulates microprocessor and links cell-density-dependent miRNA biogenesis to cancer. *Cell*. 2014;156(5):893–906.
- 46 Fang W, Bartel DP. The menu of features that define primary microRNAs and enable de novo design of microRNA genes. *Mol Cell*. 2015;60(1):131–145.
- 47 Roden C, Gaillard J, Kanoria S, et al. Novel determinants of mammalian primary microRNA processing revealed by systematic evaluation of hairpin-containing transcripts and human genetic variation. *Genome Res*. 2017;27(3):374–384.
- 48 Hecker M, Boxberger N, Illner N, et al. A genetic variant associated with multiple sclerosis inversely affects the expression of CD58 and microRNA-548ac from the same gene. *PLoS Genet*. 2019;15(2):e1007961.
- 49 Hecker M, Fitzner B, Blaschke J, Blaschke P, Zettl UK. Susceptibility variants in the CD58 gene locus point to a role of microRNA-548ac in the pathogenesis of multiple sclerosis. *Mutat Res Rev Mutat Res*. 2015;763:161–167.
- 50 Li Y, Du C, Wang W, et al. Genetic association of MiR-146a with multiple sclerosis susceptibility in the Chinese population. *Cell Physiol Biochem*. 2015;35(1):281–291.
- 51 Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015;31(21):3555–3557.
- 52 Sudmant PH, Rausch T, Gardner EJ, et al. An integrated map of structural variation in 2,504 human genomes. *Nature*. 2015;526(7571):75–81.
- 53 Kitts A, Phan L, Ward M, Holmes JB. The database of short genetic variation (dbSNP). 2013 Jun 30 [Updated 2014 Apr 3]. *The NCBI Handbook*. 2nd ed. Bethesda (MD): National Center for Biotechnology Information (US); 2013. [Internet].
- 54 Zhang Y, Zang Q, Xu B, et al. IsomiR Bank: a research resource for tracking IsomiRs. *Bioinformatics*. 2016;32(13):2069–2071.
- 55 Backes C, Fehlmann T, Kern F, et al. miRCarta: a central repository for collecting miRNA candidates. *Nucleic Acids Res*. 2018;46(D1):D160–D167.
- 56 Cloonan N, Wani S, Xu Q, et al. MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol*. 2011;12(12):R126.
- 57 Gruber AR, Bernhart SH, Lorenz R. The ViennaRNA Web Services. In: Picardi E. (ed.). *RNA Bioinformatics. Methods in Molecular Biology*, vol. 1269. New York (NY): Humana Press; 2015.
- 58 Kerpedjiev P, Hammer S, Hofacker IL. Forna (force-directed RNA): simple and effective online RNA secondary structure diagrams. *Bioinformatics*. 2015;31(20):3377–3379.
- 59 Thompson AJ, Banwell BL, Barkhof F, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol*. 2018;17(2):162–173.
- 60 Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*. 1983;33(11):1444–1452.
- 61 Van Gassen S, Gaudilliere B, Angst MS, Saeys Y, Aghaepour N. CytoNorm: a normalization algorithm for cytometry data. *Cytometry A*. 2020;97(3):268–278.
- 62 Monaco G, Chen H, Poidinger M, Chen J, de Magalhães JP, Larbi A. flowAI: automatic and interactive anomaly discerning tools for flow cytometry data. *Bioinformatics*. 2016;32(16):2473–2480.
- 63 Morbach H, Eichhorn EM, Liese JG, Girschick HJ. Reference values for B cell subpopulations from infancy to adulthood. *Clin Exp Immunol*. 2010;162(2):271–279.
- 64 Bonnefoy JY, Lecoanet-Henchoz S, Aubry JP, Gauchat JF, Graber P. CD23 and B-cell activation. *Curr Opin Immunol*. 1995;7(3):355–359.
- 65 Peltier HJ, Latham GJ. Normalization of microRNA expression levels in quantitative RT-PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues. *RNA*. 2008;14(5):844–852.
- 66 McCall MN, McMurray HR, Land H, Almudevar A. On non-detects in qPCR data. *Bioinformatics*. 2014;30(16):2310–2316.
- 67 Patsopoulos NA, Barcellos LF, Hintzen RQ, et al. Fine-mapping the genetic association of the major histocompatibility complex in multiple sclerosis: HLA and non-HLA effects. *PLoS Genet*. 2013;9(11):e1003926.
- 68 Zivković M, Stanković A, Dincić E, et al. The tag SNP for HLA-DRB1\*1501, rs3135388, is significantly associated with multiple sclerosis susceptibility: cost-effective high-throughput detection by real-time PCR. *Clin Chim Acta*. 2009;406(1–2):27–30.
- 69 Allegra D, Mertens D. *In-vivo* quantification of primary microRNA processing by Drosha with a luciferase based system. *Biochem Biophys Res Commun*. 2011;406(4):501–505. Mar 25.
- 70 Bilan V, Allegra D, Kuchenbauer F, Mertens D. *In vivo* processing assay based on a dual-luciferase reporter system to evaluate DROSHA enzymatic activity. *Methods Mol Biol*. 2014;1095:87–93.
- 71 Nguyen TA, Jo MH, Choi YG, et al. Functional anatomy of the human microprocessor. *Cell*. 2015;161(6):1374–1387.
- 72 Li S, Le TN, Nguyen TD, Trinh TA, Nguyen TA. Bulges control pri-miRNA processing in a position and strand-dependent manner. *RNA Biol*. 2021;18(11):1716–1726.

- 73 Calin GA, Ferracin M, Cimmino A, et al. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med*. 2005;353(17):1793–1801.
- 74 Tweedie S, Braschi B, Gray K, et al. Genenames.org: the HGNC and VGNC resources in 2021. *Nucleic Acids Res*. 2021;49(D1):D939–D946.
- 75 Krüger J, Rehmsmeier M. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res*. 2006;34(Web Server issue):W451–W454. Jul 1 Web Server issue.
- 76 Kinsella RJ, Kähäri A, Haider S, et al. Ensembl BioMart: a hub for data retrieval across taxonomic space. *Database (Oxf)*. 2011;2011:bar030.
- 77 Sticht C, De La Torre C, Parveen A, Gretz N. miRWalk: an online resource for prediction of microRNA binding sites. *PLoS One*. 2018;13(10):e0206239.
- 78 Huang HY, Lin YC, Li J, et al. miRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res*. 2020;48(D1):D148–D154.
- 79 Jassal B, Matthews L, Viteri G, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2020;48(D1):D498–D503.
- 80 Xie Z, Bailey A, Kuleshov MV, et al. Gene set knowledge discovery with enrich. *Curr Protoc*. 2021;1(3):e190.
- 81 UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;49(D1):D480–D489.
- 82 Novershtern N, Subramanian A, Lawton LN, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell*. 2011;144(2):296–309.
- 83 Darmanis S, Sloan SA, Zhang Y, et al. A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci U S A*. 2015;112(23):7285–7290.
- 84 Fox J, Weisberg S. *An R Companion to Applied Regression*. 3rd ed. Thousand Oaks (CA): SAGE Publications; 2018.
- 85 Bretz F, Hothorn T, Westfall P. *Multiple Comparisons Using R*. Boca Raton (FL): CRC Press; 2010.
- 86 Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;67(1):1–48.
- 87 Kim B, Jeong K, Kim VN. Genome-wide mapping of DROSHA cleavage sites on primary microRNAs and noncanonical substrates. *Mol Cell*. 2017;66(2):258–269.
- 88 Ohle LM, Ellenberger D, Flachenecker P, et al. Chances and challenges of a long-term data repository in multiple sclerosis: 20th birthday of the German MS registry. *Sci Rep*. 2021;11(1):13340.
- 89 Planas R, Jelčić I, Schippling S, Martin R, Sospedra M. Natalizumab treatment perturbs memory- and marginal zone-like B-cell homing in secondary lymphoid organs in multiple sclerosis. *Eur J Immunol*. 2012;42(3):790–798.
- 90 Traub JW, Pellkofer HL, Grondey K, et al. Natalizumab promotes activation and pro-inflammatory differentiation of peripheral B cells in multiple sclerosis patients. *J Neuroinflamm*. 2019;16(1):228.
- 91 Thompson SA, Jones JL, Cox AL, Compston DA, Coles AJ. B-cell reconstitution and BAFF after alemtuzumab (Campath-1H) treatment of multiple sclerosis. *J Clin Immunol*. 2010;30(1):99–105.
- 92 Ceronie B, Jacobs BM, Baker D, et al. Cladribine treatment of multiple sclerosis is associated with depletion of memory B cells. *J Neurol*. 2018;265(5):1199–1209.
- 93 Vistbakka J, Elovaara I, Lehtimäki T, Hagman S. Circulating microRNAs as biomarkers in progressive multiple sclerosis. *Mult Scler*. 2017;23(3):403–412.
- 94 Manna I, Iaccino E, Dattilo V, et al. Exosome-associated miRNA profile as a prognostic tool for therapy response monitoring in multiple sclerosis patients. *FASEB J*. 2018;32(8):4241–4246.
- 95 Zhang R, Tian A, Wang J, Shen X, Qi G, Tang Y. miR26a modulates Th17/T reg balance in the EAE model of multiple sclerosis by targeting IL6. *Neuromol Med*. 2015;17(1):24–34.
- 96 Honardoost MA, Kiani-Esfahani A, Ghaedi K, Etemadifar M, Salehi M. miR-326 and miR-26a, two potential markers for diagnosis of relapse and remission phases in patient with relapsing-remitting multiple sclerosis. *Gene*. 2014;544(2):128–133.
- 97 Mamelì G, Arru G, Caggiu E, et al. Natalizumab therapy modulates miR-155, miR-26a and proinflammatory cytokine expression in MS patients. *PLoS One*. 2016;11(6):e0157153.
- 98 De Felice B, Mondola P, Sasso A, et al. Small non-coding RNA signature in multiple sclerosis patients after treatment with interferon- $\beta$ . *BMC Med Genom*. 2014;7:26.
- 99 Potenza N, Mosca N, Mondola P, Damiano S, Russo A, De Felice B. Human miR-26a-5p regulates the glutamate transporter SLC1A1 (EAAT3) expression. Relevance in multiple sclerosis. *Biochim Biophys Acta Mol Basis Dis*. 2018;1864(1):317–323.
- 100 Li X, Pan X, Fu X, Yang Y, Chen J, Lin W. MicroRNA-26a: an emerging regulator of renal biology and disease. *Kidney Blood Press Res*. 2019;44(3):287–297.
- 101 Regev K, Healy BC, Paul A, et al. Identification of MS-specific serum miRNAs in an international multicenter study. *Neurol Neuroimmunol Neuroinflamm*. 2018;5(5):e491.
- 102 Yang D, Wang WZ, Zhang XM, et al. MicroRNA expression aberration in Chinese patients with relapsing remitting multiple sclerosis. *J Mol Neurosci*. 2014;52(1):131–137.
- 103 Ghadiri N, Emamnia N, Ganjalikhani-Hakemi M, et al. Analysis of the expression of mir-34a, mir-199a, mir-30c and mir-19a in peripheral blood CD4+T lymphocytes of relapsing-remitting multiple sclerosis patients. *Gene*. 2018;659:109–117.
- 104 Quintana E, Ortega FJ, Robles-Cedeño R, et al. miRNAs in cerebrospinal fluid identify patients with MS and specifically those with lipid-specific oligoclonal IgM bands. *Mult Scler*. 2017;23(13):1716–1726.
- 105 Junker A, Krumbholz M, Eisele S, et al. MicroRNA profiling of multiple sclerosis lesions identifies modulators of the regulatory protein CD47. *Brain*. 2009;132(Pt 12):3342–3352.
- 106 Gu S, Chan WY. Flexible and versatile as a chameleon-sophisticated functions of microRNA-199a. *Int J Mol Sci*. 2012;13(7):8449–8466.
- 107 Wang Q, Ye B, Wang P, Yao F, Zhang C, Yu G. Overview of microRNA-199a Regulation in Cancer. *Cancer Manag Res*. 2019;11:10327–10335.
- 108 Tan Z, Du J, Shen L, et al. miR-199a-3p affects adipocytes differentiation and fatty acid composition through targeting SCD. *Biochem Biophys Res Commun*. 2017;492(1):82–88.
- 109 Nikpay M, Beehler K, Valsesia A, et al. Genome-wide identification of circulating-miRNA expression quantitative trait loci reveals the role of several miRNAs in the regulation of cardiometabolic phenotypes. *Cardiovasc Res*. 2019;115(11):1629–1645.
- 110 Leppink J, Winston K, O'Sullivan P. Statistical significance does not imply a real effect. *Perspect Med Educ*. 2016;5(2):122–124.
- 111 Li J, Xue Y, Amin MT, et al. ncRNA-eQTL: a database to systematically evaluate the effects of SNPs on non-coding RNA expression across cancer types. *Nucleic Acids Res*. 2020;48(D1):D956–D963.
- 112 Kim K, Nguyen TD, Li S, Nguyen TA. SRSF3 recruits DROSHA to the basal junction of primary microRNAs. *RNA*. 2018;24(7):892–898.
- 113 Kim K, Baek SC, Lee YY, et al. A quantitative map of human primary microRNA processing sites. *Mol Cell*. 2021;81(16):3422–3439.
- 114 Jima DD, Zhang J, Jacobs C, et al. Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs. *Blood*. 2010;116(23):e118–e127.
- 115 Perdomo C, Campbell JD, Gerrein J, et al. MicroRNA 4423 is a primate-specific regulator of airway epithelial cell differentiation and lung carcinogenesis. *Proc Natl Acad Sci U S A*. 2013;110(47):18946–18951.
- 116 Dong Y, Simske JS. Vertebrate Claudin/PMP22/EMP22/MP20 family protein TMEM47 regulates epithelial cell junction maturation and morphogenesis. *Dev Dyn*. 2016;245(6):653–666.
- 117 Li J, Khankan RR, Caneda C, et al. Astrocyte-to-astrocyte contact and a positive feedback loop of growth factor signaling regulate astrocyte maturation. *Glia*. 2019;67(8):1571–1597.
- 118 de Rie D, Abugessaisa I, Alam T, et al. An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat Biotechnol*. 2017;35(9):872–878.
- 119 Fairfax BP, Makino S, Radhakrishnan J, et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat Genet*. 2012;44(5):502–510.
- 120 Kim-Hellmuth S, Aguet F, Oliva M, et al. Cell type-specific genetic regulation of gene expression across human tissues. *Science*. 2020;369(6509):eaaz8528.
- 121 Chen C, Ridzon DA, Broomer AJ, et al. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res*. 2005;33(20):e179.
- 122 Schamberger A, Orbán TI. 3' IsoomiR species and DNA contamination influence reliable quantification of microRNAs by stem-loop quantitative PCR. *PLoS One*. 2014;9(8):e106315.
- 123 Honda S, Kirino Y. Dumbbell-PCR: a method to quantify specific small RNA variants with a single nucleotide resolution at terminal sequences. *Nucleic Acids Res*. 2015;43(12):e77.

- 124 Faller M, Toso D, Matsunaga M, et al. DGCR8 recognizes primary transcripts of microRNAs through highly cooperative binding and formation of higher-order structures. *RNA*. 2010;16(8):1570–1583.
- 125 Mohammed EM. Environmental Influencers, MicroRNA, and Multiple Sclerosis. *J Cent Nerv Syst Dis*. 2020;12: 201179573519894955. Jan 201179573519894955.
- 126 Afrasiabi A, Fewings NL, Schibeci SD, et al. The interaction of human and *Epstein-Barr Virus* miRNAs with multiple sclerosis risk loci. *Int J Mol Sci*. 2021;22(6):2927.
- 127 Li J, Zhang Y. Current experimental strategies for intracellular target identification of microRNA. *ExRNA*. 2019;1:6.