

RESEARCH

Open Access



In silico prioritisation of microRNA-associated common variants in multiple sclerosis

Ifeolutembi A. Fashina^{1,2,3}, Claire E. McCoy^{1†} and Simon J. Furney^{4**†}

Abstract

Background Genome-wide association studies (GWAS) have highlighted over 200 autosomal variants associated with multiple sclerosis (MS). However, variants in non-coding regions such as those encoding microRNAs have not been explored thoroughly, despite strong evidence of microRNA dysregulation in MS patients and model organisms. This study explores the effect of microRNA-associated variants in MS, through the largest publicly available GWAS, which involved 47,429 MS cases and 68,374 controls.

Methods We identified SNPs within the coordinates of microRNAs, \pm 5-kb microRNA flanking regions and predicted 3'UTR target-binding sites using miRBase v22, TargetScan 7.0 RNA22 v2.0 and dbSNP v151. We established the subset of microRNA-associated SNPs which were tested in the summary statistics of the largest MS GWAS by intersecting these datasets. Next, we prioritised those microRNA-associated SNPs which are among known MS susceptibility SNPs, are in strong linkage disequilibrium with the former or meet a microRNA-specific Bonferroni-corrected threshold. Finally, we predicted the effects of those prioritised SNPs on their microRNAs and 3'UTR target-binding sites using TargetScan v7.0, miRVaS and ADmiRE.

Results We have identified 30 candidate microRNA-associated variants which meet at least one of our prioritisation criteria. Among these, we highlighted one microRNA variant rs1414273 (*MIR548AC*) and four 3'UTR microRNA-binding site variants within *SLC2A4RG* (rs6742), *CD27* (rs1059501), *MMEL1* (rs881640) and *BCL2L13* (rs2587100). We determined changes to the predicted microRNA stability and binding site recognition of these microRNA and target sites.

Conclusions We have systematically examined the functional, structural and regulatory effects of candidate MS variants among microRNAs and 3'UTR targets. This analysis allowed us to identify candidate microRNA-associated MS SNPs and highlights the value of prioritising non-coding RNA variation in GWAS. These candidate SNPs could influence microRNA regulation in MS patients. Our study is the first thorough investigation of both microRNA and 3'UTR target-binding site variation in multiple sclerosis using GWAS summary statistics.

Keywords Non-coding RNA, microRNA, 3'UTR binding sites, GWAS, microRNA–target prediction, Seed, Common variants

[†]Claire E. McCoy and Simon J. Furney are Senior Authors

*Correspondence:

Simon J. Furney
simonfurney@rcsi.ie

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Multiple sclerosis (MS) is a complex chronic neuro-inflammatory condition that affects the central nervous system (CNS). This condition leads to periods of neurological disability in patients and is the cause of most non-traumatic neurological injury in young adults [1, 2]. MS pathogenicity is linked to environmental and genetic factors [3, 4]. Understanding the genetic contributions to MS could aid in identifying candidate biomarkers and in predicting disease aetiology and progression.

Genetic studies have uncovered part of the complexity of MS. More than 13 GWAS have been carried out on MS patients and control populations since 2007 [5]. Compared to earlier study designs, GWAS have contributed the most information about MS heritability, with over 200 non-MHC variants associated with the condition since the most recent GWAS [6]. Altogether, approximately 48% of MS heritability has been accounted for [6]. However, there are challenges in interpreting the causal variants within risk loci [5, 7]. In particular, non-coding variants are not often prioritised in these interpretation strategies.

MicroRNAs (miRNAs) are small ~22nt non-coding RNAs which are well conserved across organisms. They play central roles in post-transcriptional modification by binding the 3'UTR of their targets [8] (Fig. 1A), although other interactions have been reported within the 5'UTR, coding sequences and promoter sequences of target genes [9]. Interestingly, miRNAs have been shown to be dysregulated in immune cell subsets, cerebrospinal fluid (CSF) and plasma of MS patients, as well as in the MS mouse model, experimental autoimmune encephalomyelitis (EAE) [10, 11].

Despite strong evidence of microRNA dysregulation in MS patient samples and model organisms, there is limited literature on the role of microRNA variation in MS [12–15]. In contrast, variants in microRNAs and their processing machinery have been implicated in complex conditions such as cardio-metabolic conditions,

colorectal cancer, glaucoma, Alzheimer's disease and Parkinson's disease [16–22].

We hypothesised that miRNA-associated variants are implicated in MS pathology. To explore this, we generated a bioinformatics pipeline to identify candidate MS susceptibility variants in microRNA genes and in the 3'UTR binding sites of microRNA targets using summary statistics from the most recent MS GWAS meta-analysis [6]. We then characterised and evaluated the effects of these variants using *in silico* methods and publically available datasets.

Therefore, our main objectives were collation, identification and characterisation of novel (a) microRNA gene susceptibility SNPs in MS and (b) microRNA 3'UTR binding site SNPs in MS.

Results

microRNA susceptibility SNPs

In order to investigate microRNA susceptibility SNPs in MS, we developed a prioritisation protocol as highlighted in Fig. 1. This protocol integrates common variation from dbSNP v151 with microRNA annotations from miRBase v22 [23, 24], in order to capture variation within precursor and mature microRNAs (Fig. 1B). We identified 60,638 SNPs in precursor/mature miRNA regions. These were obtained by intersecting 4573 mature and hairpin structures from miRBase with over 500 million SNPs from dbSNP v151 (Fig. 1C). This independent collation exceeds 56,911 microRNA SNPs (miR-SNPs) obtained from two older databases of microRNA variation, miRNASNPv3 and PolymiRTS [25, 26], and highlights the need to collate microRNA SNPs using more recent data.

Overall, we examined miR-SNPs which (a) are among known MS susceptibility SNPs, (b) are in linkage disequilibrium (LD) with known MS susceptibility SNPs, (c) meet a microRNA-specific adjusted Bonferroni threshold and finally (d) MS susceptibility SNPs which lie within microRNA flanking regions (Fig. 1D).

(See figure on next page.)

Fig. 1 **A** Schematic representation of microRNA transcription and microRNA–mRNA interaction. microRNAs are transcribed from DNA sequences and processed by DROSHA from the primary structure to precursor structure and by Dicer into the mature sequence. These processed mature microRNA sequences then interact with mRNA targets, leading to mRNA degradation or translational repression. **B** microRNA precursor secondary structure. Altogether, we identified SNPs which are located within precursor, mature and 5-kb microRNA flanking regions. **C** Flowchart summarising our microRNA exploration procedure. We used summary statistics from the largest MS GWAS meta-analysis [6] and two publically available datasets to investigate microRNA-associated variation in MS. To capture variation within human microRNAs, we extracted the genomic coordinates of human microRNA precursor and mature regions from miRBase v22 and intersected these with all human variants recorded in dbSNP v151. In addition, we extended the precursor regions by 5 kb up- and downstream to incorporate SNPs within regulatory features (**D**). Overall, among the SNPs tested in the IMSCG's meta-analysis, we identified 314 SNPs within microRNA precursor/mature regions and 36,841 SNPs in 5-kb flanking regions. **D** In our prioritisation process, we identified microRNA SNPs (1) among known MS susceptibility SNPs, (2) in strong Linkage Disequilibrium (LD) with known MS risk SNPs and (3) which meet the adjusted *p* value threshold for the 314 microRNA SNPs tested in the IMSCG meta-analysis. **A** and **D** adapted from "microRNA in Cancer" and "The Principle of a Genome-wide Association Study (GWAS)" in Biorender.com (2022). Retrieved from <https://app.biorender.com/biorender-templates>

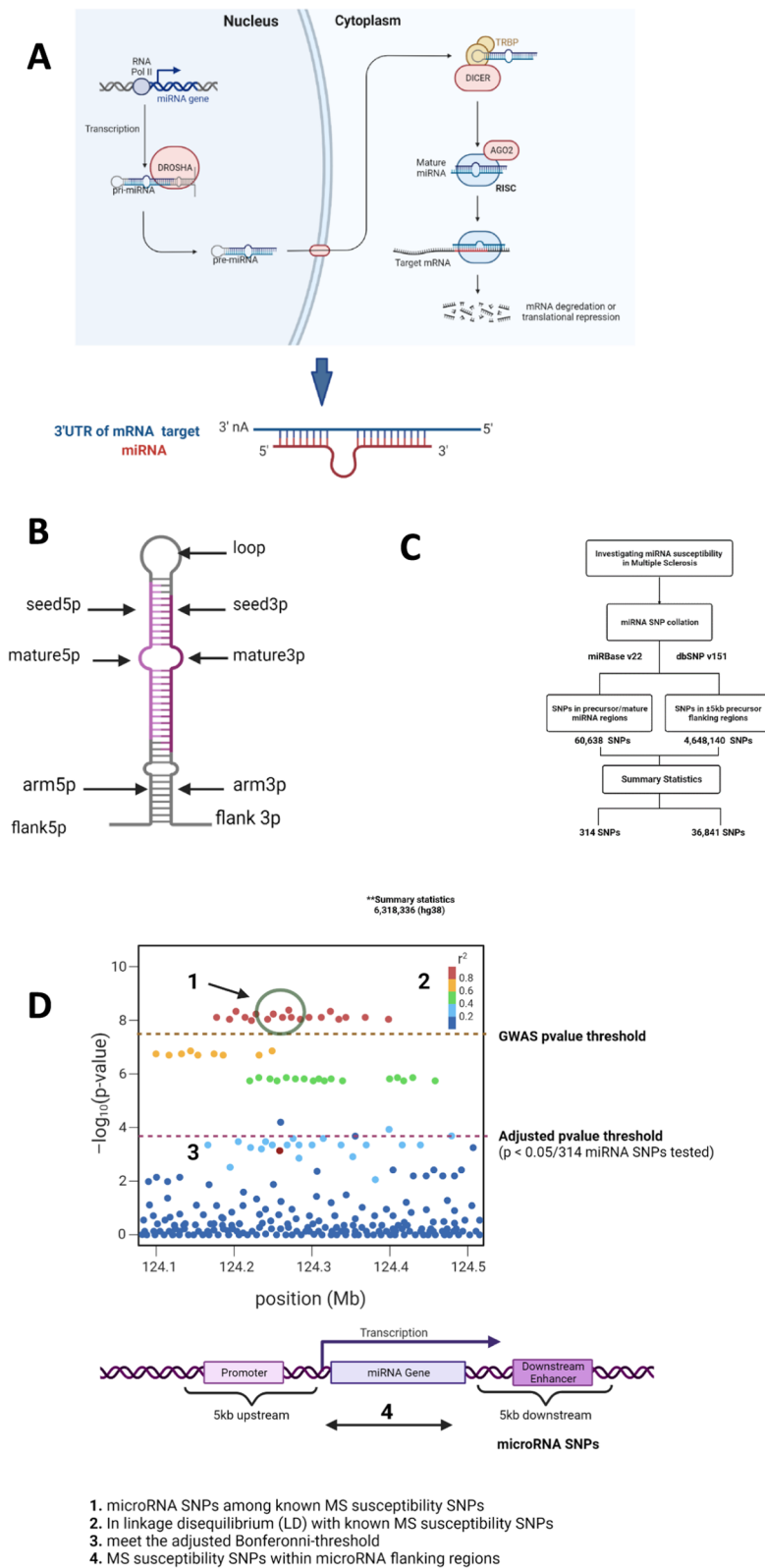


Fig. 1 (See legend on previous page.)

MicroRNA variants among known MS SNPs

Initially, we identified 314 SNPs within precursor/mature microRNAs that were tested within the 2019 GWAS summary statistics [6] (Additional file 1: Table S1). However, none of these microRNA variants were among the reported MS susceptibility SNPs. To investigate this further, we mapped miR-SNPs which were in linkage disequilibrium (LD) with the susceptibility SNPs, in order to capture tagging miR-SNPs within the MS susceptibility loci. 3 of the 314 miR-SNPs (rs1414273, rs7247237 and rs7247767) were in LD with the susceptibility SNPs (see Methods). Among these 3 miR-SNPs, rs1414273 is in high LD ($r^2=0.97$) with known MS susceptibility SNP rs10801908 (*CD58*), meets the genome-wide p value threshold ($p=8.48 \times 10^{-16}$) and lies within the 3' end of precursor hsa-mir-548ac (Fig. 2A). Only recently has attention been drawn to rs1414273/*MIR548AC* in MS. Hecker et al. [14] found that rs1414273 (*MIR548AC*) decouples transcription of *CD58* and *MIR548AC*. Their findings support our independent prioritisation process for non-coding candidate variants which are in high LD with coding MS susceptibility SNPs.

Next, we examined the effect of rs1414273 on the structural conformation of hsa-mir-548ac using miRVaS [27]. This SNP significantly impacts the flank and arm of the microRNA in both the 5' and 3' directions. The fold energy changes across the centroid (CEN), minimum fold energy (MFE) and maximum estimate accuracy (MEA) models are presented in Additional file 1: Table S2. We have represented the MEA, which has a net change of -4.4Δ in Fig. 2B and C.

In Fig. 2C, the risk allele (C) is predicted to create a more thermodynamically stable RNA secondary structure compared to the T allele. This could allow for more microRNA regulatory activity. In conclusion, rs1414273 is potentially associated with MS susceptibility and changes to *MIR548AC* stability.

Genome-wide microRNA variants

Next, to specifically focus on miRNA-related variants, we carried out Bonferroni correction on the p value threshold, adjusting for the 314 miRNA SNPs that were tested in the summary statistics. This process yielded 6 candidate MS-associated miR-SNPs within the precursor,

loop and seed regions of 4 microRNAs: *MIR548AC*, *MIR1208*, *MIR3135b* and *MIR6891* (Table 1). We established the functional implication of rs1414273 in *MIR548AC* in the previous step; therefore, we focused on the other 5 SNPs here. We examined the genomic context of the miR-SNPs and compared our candidates to reported GWAS signals.rs2648841 in the 3' end of precursor *MIR1208* (Table 1), which represents a signal separate from the IMSCG GWAS signals rs735542 (hg37 chr8:128175696), rs6990534 (*PVT*) (Fig. 2D), the proximal peaks rs11989574 and rs759648 (chr8:129158945 in another European GWAS) (Fig. 2E). Similar to rs1414273 (*MIR548AC*) above, rs2648841 was not among the prioritised effects and therefore was not among suggestive, non-replicated or no data effects outlined by the IMSCG. However, we dropped this SNP due to its nominal p value compared to the lead SNPs and because no structural changes were predicted in 2 of 3 thermodynamic models (Additional file 2: Fig. S1).

Also among the 6 miR-SNPs (Table 1), 3 are within hsa-miR-6891-3p, a product of *MIR6891* (rs17881225, rs2276448 and rs2854001). This microRNA is encoded within intron 4 of *HLA-B* and is co-transcribed with the mRNA, which is itself associated with MS susceptibility. Although these 3 candidate SNPs could have effects on target regulation, the GWAS signal is likely coming from the HLA variants identified by the IMSCG (rs2308655, rs3819284, rs1050556, *HLA-B*52:01*, *HLA-B*38:01* and *HLA-B*35:03*). Among these 3 *MIR6891* SNPs, rs2276448 lies within the seed region of the microRNA and possibly has the most significant effect on its target regulatory function compared to the SNPs in the mature and precursor loop regions. We explored the target-binding consequences of the seed SNP rs2276448 (*MIR6891*) in Additional file text (Additional file 2: Fig. S2).

Finally, rs4285314 lies in the precursor 3' end of *MIR3135b*, but is within the same susceptibility locus as the *HLA-B* variants, presenting the same challenge as the *MIR6891* variants.

Overall, having investigated the SNP in *MIR548AC*, we did not further prioritise any of the 5 Bonferroni-adjusted microRNA SNPs due to a) lack of predicted structural changes or b) high linkage disequilibrium in the MHC locus.

(See figure on next page.)

Fig. 2 A Regional LocusZoom plot [62] showing high Linkage Disequilibrium ($r^2 > 0.9$) between known MS susceptibility SNP rs10801908 (*CD58*) and our candidate SNP rs1414273, which lies in *MIR548AC*. Next, we highlight the predicted RNA secondary structure of the **B** reference sequence of hsa-mir-548ac compared to the **C** alternative (risk) allele. This figure shows the MEA prediction which has the greatest net change in free energy among the 3 models predicted by miRVaS (Additional file 1: Table S2). The arrow highlights the SNP rs1414273 (in red) located in the 3' end (arm) of precursor sequence. Lower free energy measures indicate greater RNA stability; therefore, microRNA with the alternative risk allele is more thermodynamically stable than the reference allele. Candidate SNP rs2648841 is within genomic coordinates of *MIR1208*. **D** This variant represents a different signal from IMSCG SNPs rs6990534 and rs735542 (chr8:128175696) and **E** is not in LD with rs11989574, the peak SNP in its genomic region or rs1861842 (not shown) and rs759648 (chr8:129158945) which were implicated in African Americans and Europeans, respectively [53]. Although this SNP is below genome-wide significance ($p = 3.86 \times 10^{-5}$), its association with MS cannot be ruled out

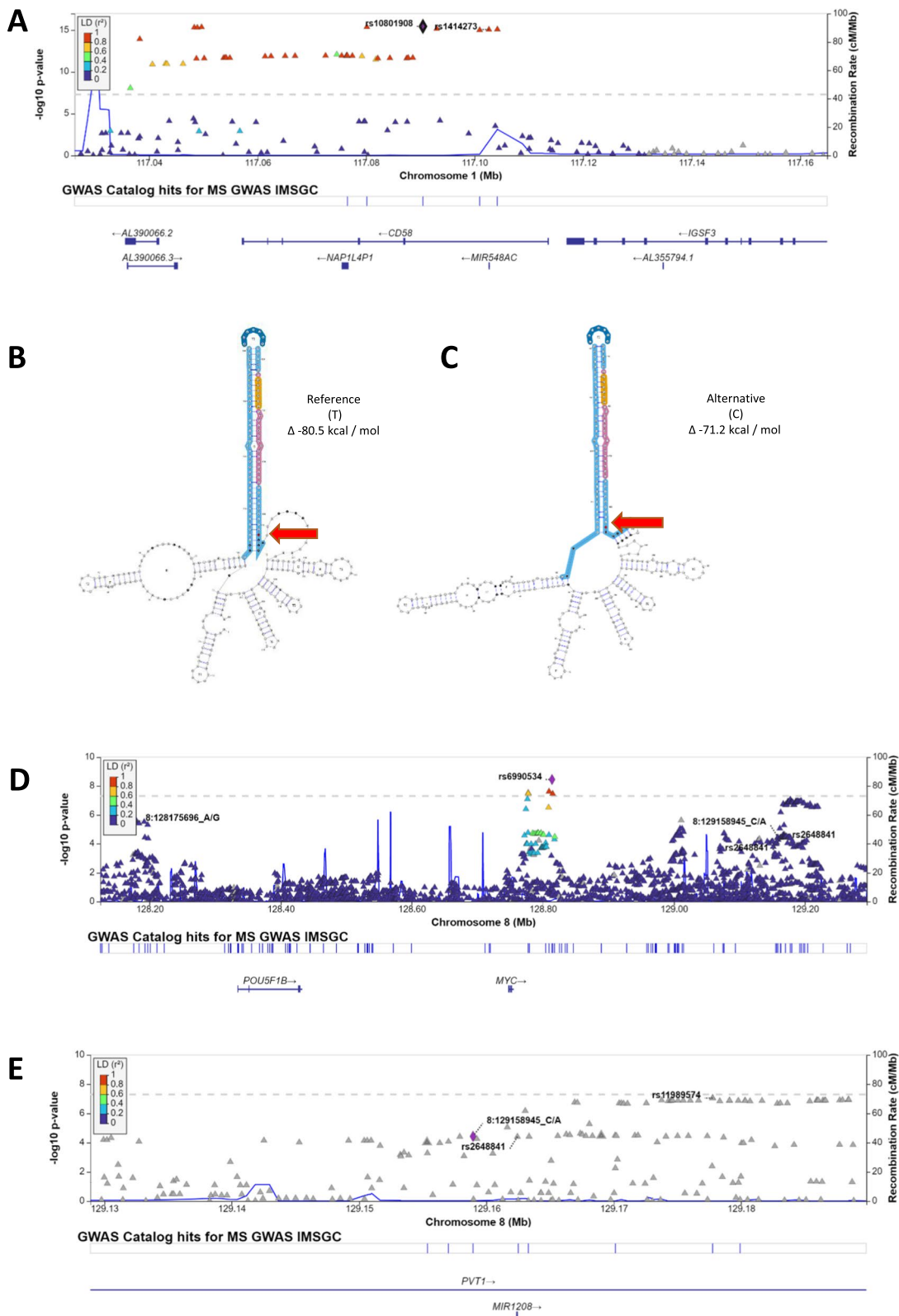


Fig. 2 (See legend on previous page.)

Table 1 Annotation of the 6 miRNA SNPs that passed the Bonferroni-adjusted p value threshold ($p < 0.05/314$)

| CHR: POS (hg38) | RSID | A1/A2 | EUR_AF | Discovery GWAS p value | OR | miRNA | ADmiRE annotations | OR interpretation | Prioritisation results |
|-----------------|------------|-------|--------|--------------------------|--------|-----------|---------------------|-------------------|------------------------------|
| 1:116560027 | rs1414273 | C/T | 0.1402 | 8.48E−16 | 1.2333 | mir-548ac | Precursor_3PrimeEnd | risk | High LD, structural change |
| 8:128150187 | rs2648841 | G/A,T | 0.0179 | 3.86E−05 | 1.1322 | mir-1208 | Precursor_3PrimeEnd | risk | No structural change |
| 6:31355288 | rs17881225 | G/C | 0.0984 | 7.72E−05 | 1.2163 | mir-6891 | Precursor_Loop | risk | Signal in high LD HLA region |
| 6:31355243 | rs2276448 | T/C | 0.2366 | 7.73E−12 | 1.2517 | mir-6891 | 3p_Seed | risk | Signal in high LD HLA region |
| 6:31355235 | rs2854001 | G/A | 0.2117 | 5.44E−38 | 1.5888 | mir-6891 | 3p_Mature | risk | Signal in high LD HLA region |
| 6:32749925 | rs4285314 | G/A | 0.5318 | 7.1E−122 | 1.5134 | mir-3135b | Precursor_3PrimeEnd | risk | Signal in high LD HLA region |

ADmiRE [28] was used to identify the locations of these SNPs. In order of consequence, SNPs in the seed region > mature > loop > precursor ends. The association with MIR548AC was explored in the previous section. The structural consequence of MIR1208 SNP was explored, while the microRNA-binding ability of MIR6891 was examined in the context of the seed SNP. Discovery GWAS p values and ORs of these SNPs are also presented in context

microRNA flanking SNPs among known MS SNPs

Having explored our microRNA variants with regard to susceptibility SNPs, LD and adjusted p values, we expanded our definition of microRNA variants to include those within ± 5 -kb flanking regions of the precursors, in a similar approach to that employed by Fang and colleagues [21]. By extending the miRNA precursor coordinates by ± 5 kb, we aimed to incorporate microRNA regulatory features that might be influencing their expression. We found over 4 million SNPs in ± 5 -kb precursor flanking regions of miRNAs (Fig. 1C). In total, 36,841 of these were tested in the summary statistics (Fig. 1C). Among these, two variants proximal to hsa-mir-10399 and hsa-mir-4492 were among known susceptibility SNPs. rs10271373 and rs149114341 (chr11:118783424 hg37) are downstream of hsa-mir-10399 and hsa-mir-4492, respectively (Additional file 1: Table S5), and were annotated as intergenic SNPs in the 2019 GWAS meta-analysis. An enhancer sequence

is reported around the coordinates of rs149114341. However, we were unable to characterise its effects on MIR4492 using summary statistics only. rs10271373 maps downstream of MIR10399 as well as to the 3'UTR binding site of ZC3HAV1, a gene that has been implicated in MS [6]. Therefore, this SNP was prioritised in our 3'UTR binding site analysis instead.

Altogether, our prioritisation process highlights rs1414273 (MIR548AC) as a candidate MS SNP among the other candidates (Additional file 1: Table S10).

3'UTR microRNA-binding site susceptibility SNPs

Variants in the 3'UTRs of mRNAs could disrupt or create microRNA-binding sites, contributing to transcriptional dysregulation. We explored variation in 3'UTR microRNA-binding sites that could be relevant to MS, by implementing a procedure similar to our microRNA variant collation (Fig. 3A).

(See figure on next page.)

Fig. 3 A Flowchart showing our 3' UTR microRNA-binding site exploration pipeline using 3 publicly available datasets and summary statistics provided by IMSCG (2019). 3'UTR variant collation was performed separately from microRNA variant collation. We obtained variants from dbSNP v151 and integrated these into the 3'UTR binding sites predicted by RNA22 v2.0 and TargetScan v7.0 (see Methods) **(B)**. Schematic showing the predicted effects of rs6742 on microRNA-binding ability to 3'UTR in SLC2A4RG. The C allele is expected to bind to 6 microRNAs differently to the T allele. Overall, we expect that the C allele is under stronger regulation than the T allele. **C** Venn Diagram showing the overlap between GWAS independent SNPs and our collection of 3'UTR SNPs which are in 3'UTR binding sites. Independent SNPs were identified through FUMA and a list of suggestive effects provided by the IMSCG. **D** Among the 19 independent SNPs from **C**, we tested the microRNA-binding ability of the 3'UTR binding sites containing 8 SNPs. Among these, 6 SNPs were found to cause microRNA-binding site changes in their 3'UTR sites. Here, we show the number of microRNAs that bind to the alternative and reference versions of the 3'UTR sequences, as well as the microRNAs that bind differently. The source column highlights which GWAS independent list the SNP has been output from. **E** LocusZoom regional plot showing the 3'UTR SNP rs2587100, which is independent, weakly suggestive, causes changes in microRNA-binding ability of BCL2L13 and is an eQTL for BCL2L13 in general monocytes and MS patient monocytes. This is our only candidate SNP which has MS patient specific eQTL evidence. No other SNPs in this region were prioritised among the genome-wide IMSCG SNPs. The highlighted intronic SNP rs9618043 (CECR2) is among the non-replicated SNPs (NR) from the IMSCG's prioritised effects within this region (IMSCG Additional file 1: Table S6), while rs9618040 is not among the prioritised effects (intron CECR2)

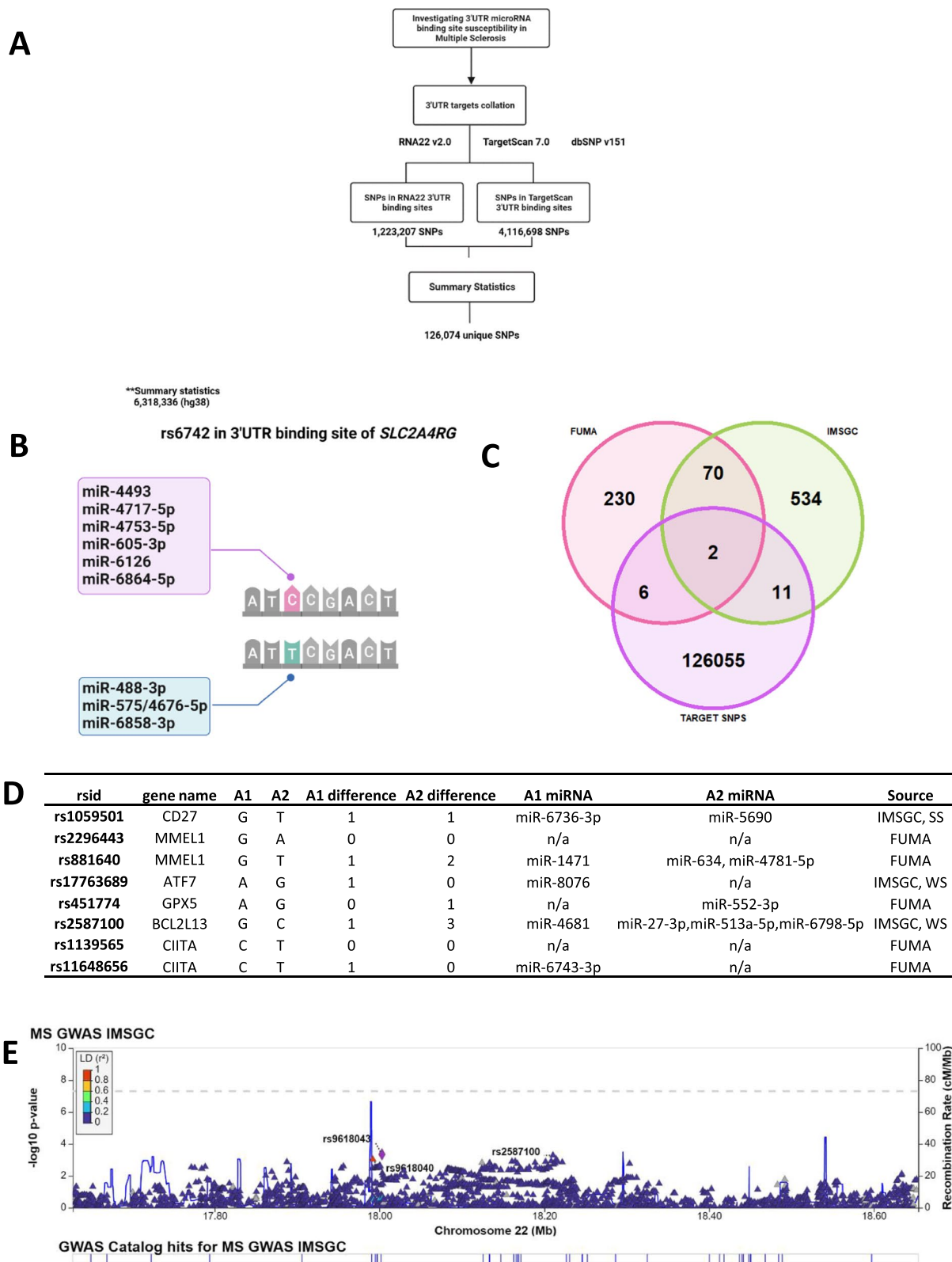


Fig. 3 (See legend on previous page.)

3'UTR microRNA-binding sites were extracted from two microRNA–target prediction tools which implement different algorithms. Predictions from both TargetScan v7.0 and RNA22 v2.0 were used to capture microRNA–target interactions within the 3'UTRs [29, 30]. Among over 14 million RNA22-predicted binding sites, 1,223,207 sites were retained as they had the most significant p values per miRNA–target pair. In addition, over 15 million TargetScan-predicted binding sites were identified. Together, we found 1,223,207 RNA22 SNPs and 4,116,698 TargetScan SNPs after intersecting dbSNP v151 with the genomic coordinates of these 3'UTR binding sites (Fig. 3A). We collated 126,074 SNPs among the predicted 3'UTR binding sites which were also tested in the summary statistics.

3'UTR variants among known MS SNPs

Overall, we identified two [2] IMSGC susceptibility SNPs among our collated 3'UTR binding site variants. rs10271373 ($p=3.11 \times 10^{-9}$, GWAS joint OR=0.946) and rs6742 ($p=4.11 \times 10^{-14}$, GWAS joint OR=1.149) lie within the predicted 3'UTR microRNA-binding sites of *ZC3HAV1* and *SLC2A4RG*, respectively. We used TargetScan v7.0 to analyse miRNA-binding changes in reference versus variant 3'UTR sequences for these 2 candidate mRNAs. Of the 2 susceptibility SNPs, we observed changes in miRNA-binding ability for rs6742 only. In short, rs6742 changes which miRNAs can bind to the *SLC2A4RG* 3'UTR of both alleles (Additional file 1: Table S6, Fig. 3B). Overall, the risk allele of rs6742 appears to be under tighter miRNA regulation than the T allele, with a net change of +3 miRNA interactions (Fig. 3B).

3'UTR variants among independent SNPs

We postulated that more candidate 3'UTR binding site MS SNPs exist, but were not prioritised as susceptibility SNPs in the 2019 meta-analysis. Therefore, to broaden our 3'UTR binding site candidates, we first identified independent SNPs from the IMSGC's Additional file [6] (Additional file 1: Table S11). In total, we extracted a list of 201 independent genome-wide SNPs and 416 independent weakly and strongly suggestive (299 WS and 117 SS) SNPs. For some of these suggestive SNPs, their joint p values were greater than their discovery p values; however, they did not meet genome-wide significance, while others replicated significantly in only one dataset [6] (Additional file 1: Table S11). We then explored whether any of our 126,074 3'UTR SNPs, which were tested in the summary statistics were among these independent SNPs from IMSGC. On intersecting both datasets, we identified 13 3'UTR SNPs within the IMSGC's independent SNPs (Additional file 1: Table S7, Fig. 3C).

To expand the methodology used to identify independent SNPs within the summary statistics, we uploaded the IMSGC GWAS summary statistics into FUMA Webtools [31]. While the IMSGC applied stepwise conditional regression to their discovery and replication cohorts to identify independent effects, FUMA uses PLINK's [32] clumping procedure to rank independent and lead SNPs from GWAS summary statistics. We identified 318 independent SNPs from the FUMA web tools. Eight of our 3'UTR SNPs were among the FUMA independent SNPs (Additional file 1: Table S7, Fig. 3C). Additionally, two [2] 3'UTR independent SNPs were shared by both FUMA and the IMSGC (Fig. 3C). Altogether, we identified 19 3'UTR SNPs among the independent SNPs (Additional file 1: Table S7).

We set out to investigate these 19 3'UTR SNPs through *in silico* methods and publically available functional evidence (see Methods). Other studies [33–35] have proposed criteria to validate microRNA–target interactions. These can be summarised as (1) demonstration of co-expression, (2) direct interaction between miRNA and region on target, (3) gain and loss experiments to show target protein interaction and (4) predicted changes have biological functions. We incorporated these approaches into our prioritisation process (see Methods). In short, functionally relevant 3'UTR SNPs are likely to change miRNA–target interactions at the 3'UTR binding site, act as eQTLs for the targets in MS relevant tissues (e.g. PBMCs, lymphocytes) and have the relevant microRNAs expressed in the same MS relevant tissues. We investigated these using the FiveX browser for eQTL catalogue, an MS eQTL dataset from the IMSGC, RegulomeDB v2.07 and FANTOM5 [6, 36–38] (Additional file 1: Table S11). These criteria are highlighted in Additional file 2: Fig. S3.

Among our 19 candidates, we excluded 11 for the following reasons. Two had been assessed in the susceptibility SNP process, 7 had been reannotated as intronic and the relevant 3'UTR sequences were unavailable for 2 SNPs. Therefore, we carried out microRNA gain/loss on 8 independent SNPs using TargetScan 7.0 (Additional file 1: Table S7). Among these 8, we found that 6 SNPs (rs1059501 *CD27*, rs881640 *MMEL1*, rs2587100 *BCL2L13*, rs11648656 *CIITA*, rs17763689 *ATF7* and rs451774 *GPX5*) change the microRNA-binding ability of 3'UTRs (Fig. 3D, Additional file 1: Table S8).

The 3'UTR of *BCL2L13* has the greatest changes in microRNA binding due to rs2587100 (GWAS joint OR = 3.43×10^{-05} , joint p value = 1.049). The G allele (risk allele) binds to miR-4681 while the C allele binds to miR-27-3p, miR-513a-5p and miR-6798-5p, suggesting that the 3'UTR sequence which includes the risk allele is possibly under less regulation (Fig. 3D).

rs2587100 was among the weakly suggestive effects (IMSGC), but has strong functional support for its potential role. rs2587100 is also the only SNP among our 6 candidates which is an eQTL for the target gene in non-MS and an MS patient dataset (see Methods). However, the effect of the eQTL is not consistent across the non-MS and 1 MS dataset. It decreases *BCL2L13* expression in non-MS PBMCs and monocytes, and increases *BCL2L13* expression in MS PBMCs and monocytes, and its relevant microRNAs are also expressed in monocytes [6, 38–40] (Additional file 1: Table S9, Additional file 1: Table S11). We have presented the genomic context of this SNP in Fig. 3E. Apart from rs9618043 *CECR2*, which is among the IMSGC's non-replicated SNPs, no other SNPs in this region were among the prioritised genome-wide IMSGC SNPs.

Only two [2] other independent 3'UTR SNPs (rs1059501 and rs881640) met the functional validation criteria (Fig. 4A, Additional file 1: Table S9). The significant eQTL activity or microRNA expression for rs11648656, rs17763689 and rs451774 appears to not be relevant for MS tissues (Additional file 1: Table S9, Additional file 2: Fig. S4). Therefore, we prioritised rs1059501 and rs881640 and have shown their genomic context in Fig. 4B and C. rs1059501 (*CD27*) is independent from the IMSGC susceptibility SNPs (rs1800693, rs2364485 and rs12832171) in that region and was ranked as strongly suggestive in the IMSGC stepwise regression. Both the protective (*G*) and alternative allele (*T*) lose and gain one miRNA, respectively (Fig. 3D). The microRNAs gained/lost due to this SNP are expressed in monocytes and haematopoietic cells (Additional file 1: Table S9) [38], while the SNP has been shown to decrease *CD27* expression in T-cells and LCLs, and increase *CD27* in monocytes and brain tissue [41–45]. Finally, rs881640 is independent from the IMSGC genome-wide SNP [chr1:2520527(hg37); rs6670198] in that region. Its *G* (risk) allele binds to miR-1471, but not to miR-634 or miR-4781, which are recognised by the *T* allele. Therefore, we expect that 3'UTR sequence containing the *G* allele is likely under less regulation than the *T* allele. This SNP has been shown to decrease *MMEL1* expression in blood, monocytes and T-cells [41, 46, 47].

Overall, we highlighted that the impact of increased regulation of *SLC2A4RG* due to the new miRNA interactions could be significant. In addition, 3 independent SNPs in 3'UTRs of *BCL2L13*, *CD27* and *MMEL1* meet multiple miRNA:target interaction criteria. Therefore, we have presented evidence that these SNPs could be involved in MS pathogenesis and should be prioritised for future investigation.

Discussion

In this study, we presented evidence that microRNA-associated variants could be implicated in MS. Our analysis is the first systematic exploration of both microRNA and 3'UTR target-binding site variation in MS, using GWAS summary statistics. By using the most recent meta-analysis [6], we harnessed the largest MS GWAS resource available to test our hypothesis. Altogether, we identified 30 candidate microRNA-associated variants from our collation procedure. Those variants meet a microRNA-specific Bonferroni-corrected threshold, are in LD (Linkage Disequilibrium) with known susceptibility SNPs or are suggestive SNPs from the IMSGC GWAS [6], whose microRNA functions had not been evaluated previously. We prioritised 1 of 8 miR-SNPs, rs1414273 (*MIR548AC*), and 4 of 22 SNPs in 3'UTR microRNA-binding sites of *SLC2A4RG* (rs6742), *CD27* (rs1059501), *MMEL1* (rs881640) and *BCL2L13* (rs2587100), based on structural and functional predictions. Therefore, these 5 SNPs are our top candidate microRNA-associated variants which could play a role in MS pathogenesis.

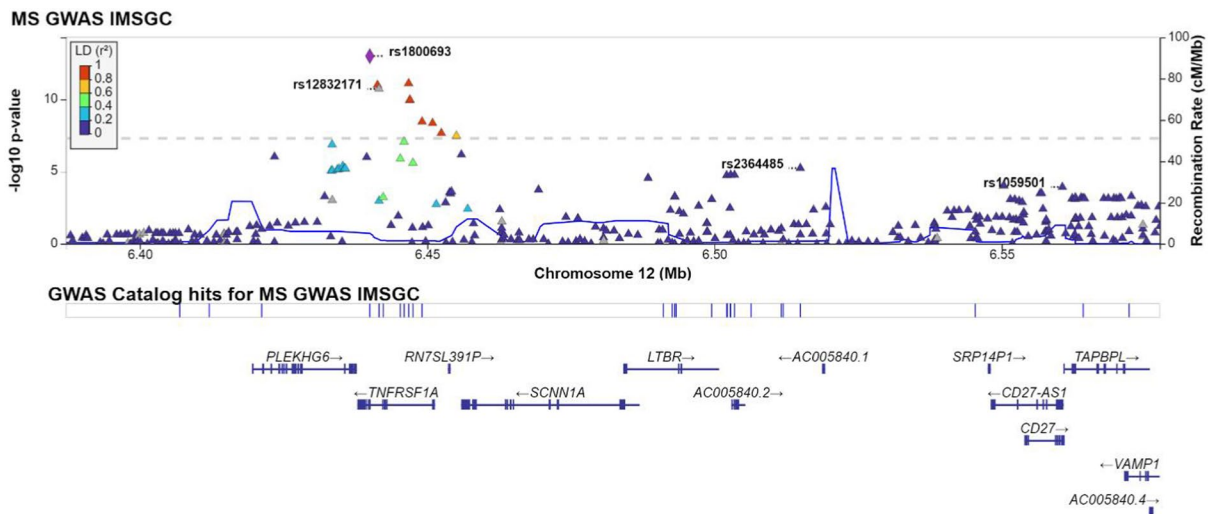
Our work successfully incorporates multiple microRNA prioritisation methods used elsewhere [12, 19, 21, 22]. The most relevant comparison to our results is a study which implicated 11 microRNAs in MS susceptibility [48]. We noted one major difference in methodology. Hecker and colleagues [48] analysed all SNPs within microRNA stem loop and Drosha cleavage sites which are close (<250 kb) to the 233 IMSGC GWAS SNPs, irrespective of their presence in the summary statistics. Therefore, association analysis-based *p* values were not factored in within their prioritisation process. Among of their 12 candidate SNPs, 6 had been tested in the GWAS and were not Bonferroni-corrected. However, both studies identified hsa-mir-548ac and hsa-mir-4492 as candidate MS microRNAs. Another important comparison is against a microRNA GWAS study performed on a paediatric MS cohort [12]. Rhead et al. [12] adjusted the *p* value threshold for microRNA variants within a paediatric cohort of MS patients, but did not identify any significant SNPs. To follow up, those authors used MIGWAS [49] to identify enrichment of candidate microRNA–target network signals. Alternatively, we examined our candidates individually, to characterise effects of the variants on the functions of microRNAs and targets directly.

Parallel to other studies, we successfully examined the effect of risk alleles through *in silico* methods [22, 50–52]. Our secondary structure prediction spotlighted that the risk allele for rs1414273 is expected to yield higher *MIR548AC* levels. Interestingly, rs1414273 has been shown to decouple the transcription of miR-548ac from its host gene *CD58*, leading to increased levels of miR-548ac [14]. This is in line with our secondary

A

| prioritised SNPs | gene name | A1 | A2 | A1 diff | A2 diff | joint OR/ discovery OR where joint not available | joint P/discovery P where joint not available | risk interpretation |
|------------------|----------------|----|----|---------|---------|--|---|---------------------|
| rs1059501 | CD27 | G | T | 1 | 1 | 0.944 | 6.38E-07 | protective |
| rs881640 | MMEL1 | G | T | 1 | 2 | 1.1782 | 4.33E-10 | risk |
| rs2587100 | BCL2L13 | G | C | 1 | 3 | 1.049 | 3.43E-05 | risk |
| rs11648656 | CIITA | C | T | 1 | 0 | 0.8672 | 6.86E-13 | protective |
| rs17763689 | ATF7 | A | G | 1 | 0 | 0.913 | 6.94E-04 | protective |
| rs451774 | GPX5 | A | G | 0 | 1 | 0.8773 | 1.42E-13 | protective |

B



C

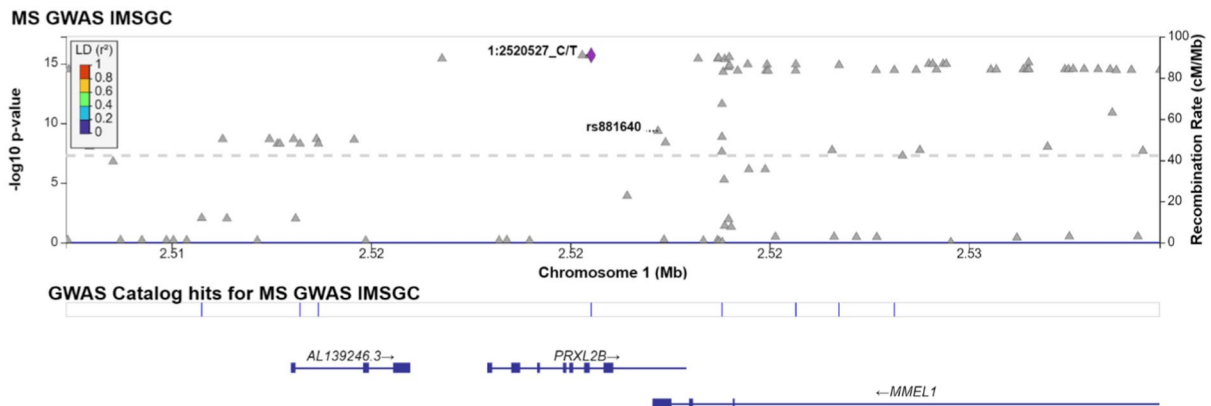


Fig. 4 **A** Table showing the p values and odds ratio (OR) for the 6 independent 3'UTR independent SNPs. Among these, only 3 (highlighted in bold) meet 3 of the microRNA–target validation criteria (see Methods, Additional file 1: Table S9). Joint p values (from IMSGC's discovery and replication processes) are available for the IMSGC independent (suggestive) SNPs, but not for those identified by FUMA, as these were not among the suggestive effects. For the latter group, we have showed the discovery p values and ORs. **B** LocusZoom plots showing regions around our other 2 functionally relevant SNPs (rs2587100 is in Fig. 3E). We have highlighted the 3'UTR SNPs in rs1059501 (*CD27*) and rs881640 (*MMEL1*). Our candidate SNP rs1059501 is independent from the IMSGC susceptibility/genome-wide SNPs (rs1800693, rs2364485, rs12832171) in that region and was ranked as strongly suggestive in the IMSGC stepwise regression. **C** Our candidate SNP rs881640 is independent from the IMSGC susceptibility/genome-wide SNP (chr1:2520527(hg37); rs6670198) in that region

stability model. Despite rs1414273 [chr1:117102649 (hg37), 0.14 EUR MAF] being significant in the discovery cohort of the IMSGC meta-analysis, it was not among the effects prioritised for replication. It also appears not to have been captured among the 46 SNPs within that haplotype between the two replication datasets [6] (Additional file 1: Table S11). Conversely, the IMSGC's prioritised effect SNP rs10801908 alone might paint an incomplete picture, due to the presence of *MIR548AC* within the first intron of *CD58* and the strong linkage between rs10801908 and rs1414273. However, there is limited research into the role of *MIR548AC* in immunological conditions. Next, although our candidate rs2648841 did not change the structural conformation of miR-1208, another *MIR1208* SNP rs1861842 has been associated with MS in African Americans [53], implicating the microRNA further.

In additional *in silico* experiments, we identified changes to *MIR6891*'s binding ability, which could lead to functional changes to the mRNA. However, because *MIR6891* lies within an intron of *HLA-B*, a class I MHC molecule with protective MS SNPs [54, 55], it is challenging to segregate the MHC signal from the microRNA signal using only summary statistics. While *MIR6891* seed SNP rs2276448 itself has not been assessed in MS, miR-6891-3p is linked to changes in macrophage-driven inflammation [56]. Our microRNA gain/loss analysis also showed that the risk allele of rs6742 in *SLC2A4RG* is likely under stronger microRNA regulation than the other allele. This is supported by the IMSGC's 2019 study, where SNPs within the rs6742 susceptibility locus were all associated with reduced *SLC2A4RG* expression in CD4+ T-cells in an MS cohort [6] (Additional file 1: Table S11). *SLC2A4RG* functions as a transcription factor for *SLC2A/GLUT4*, which is among the glucose carriers that are upregulated following lymphocyte activation [57]. This highlights a possible link between *SLC2A4RG* dysregulation in CD4+ T-cells and T-cell activation.

After broadening our search for independent SNPs through FUMA, we identified changes to the microRNA-binding ability of *CD27*, *MMEL1* and *BCL2L13* due to rs1059501, rs881640 and rs2587100, respectively. This highlights the value of using different methods to identify independent SNPs. The eQTL rs2587100 drives increased expression of *BCL2L13* in MS patients [6] (Additional file 1: Table S11) and aligns with our microRNA gain/loss experiment which shows that the risk allele is under less regulation than the C allele. *BCL2L13* has been linked to mitophagy [58]; therefore, investigation of this upregulation in monocytes could be important. However, surprisingly, the non-MS eQTLs for both *BCL2L13* and *MMEL1* reduce their expression [38–40]. Genotyping these SNPs

directly in MS patients could clarify the true direction of this eQTL.

Next, at least one other group has incorporated flanking regions in microRNA-specific GWAS, in order to explore regulatory features which may influence microRNA transcription [21]. Our identification of a risk SNP in an enhancer-like domain [59] flanking *MIR4492* suggests the regulation of these microRNA genes by other factors. Expression of this microRNA within B-cells is proposed to be altered due to Epstein–Barr Virus (EBV) infection, which has been shown to increase MS risk significantly [4, 13]. The effect of this enhancer SNP on *MIR4492* expression in MS patients should be investigated further, especially in the context of EBV infection.

The main challenges with interpreting our findings are the long-range LD in the MHC region, limited microRNA annotations and the ability of microRNAs to bind to multiple targets. We identified consequences of seed SNP rs2276448 (*MIR6891*), but could not confirm its independence from HLA-B SNPs (rs2308655, rs3819284, rs1050556, *HLA-B*52:01*, *HLA-B*38:01* and *HLA-B*35:03*) using only publically available data. Furthermore, we were unable to measure the effect of multiple candidate SNPs on microRNAs or their targets by using only summary statistics. We also could not annotate the flanking SNPs which exceeded the 2-kb region stipulated by miRVaS. This is a challenge with microRNA tools such as miRVaS, as promoter information is not often available for intergenic microRNAs [60]. This means that the microRNA mapping tools are not fully powered to identify SNPs in enhancer regions, transcription start sites, among others; therefore, this needs to be accounted for in downstream analysis. Finally, experimental validation of our predicted changes in *MIR6891*, *SLC2A4RG*, *CD27*, *MMEL1* and *BCL2L13* will be necessary in the future due to the limitations of microRNA–target prediction algorithms. Finally, this study was limited to publically available data; therefore, the eQTL data were sourced from multiple studies.

Conclusions

Altogether, we identified 30 candidate microRNA-associated variants through systematic analysis of MS GWAS summary statistics. We prioritised 1 microRNA SNP and 4 3'UTR binding site SNPs based on the effects of the MS variants on their function, structure or regulatory abilities. Our *in silico* work helps to bridge the gap between MS GWAS and microRNAs implicated in MS.

Methods

Summary statistics

Summary statistics from the most recent GWAS meta-analysis [6] on MS patients were requested from the

IMSGC through the webpage (<https://imsgc.net/>). In short, over 8 million SNPs were imputed and tested for 47,429 MS cases and 68,374 control subjects by the consortium.

Genomic coordinates for all summary statistics (including autosomal and non-autosomal SNPs) were provided in hg37. We lifted over to hg38 using Ensembl's [61] Assembly Converter for downstream hg38 SNP integration. We visualised all regional associations in LocusZoom's web platform [62]. An overview of the pipeline and tools is presented in Additional file 2: Fig. S5.

Text mining

Prior to collating microRNA SNPs, we wanted to test whether the microRNA-associated variant databases PolymiRTS [26] and miRNASNP v3 [25] were up to date. miRNASNP v3 contains SNPs in microRNA seed and precursor regions, target 3'UTR SNPs as well as predictions of miRNA gain/loss based on these 3'UTR SNPs. PolymiRTS was last updated in 2014 and contains microRNA seed regions from miRBase v20 and 3'UTR sequences for CLASH validated targets. Altogether, this resulted in the collation of 56,911 SNPs. We compared microRNA SNPs from the literature to those within the databases. Specifically, the term "microRNA" was used in PubMed's eFetch commandline tool, to obtain abstracts for all relevant papers published between 2014 and 2021. We then extracted rsids from these abstracts and manually confirmed whether the SNPs were referring to the microRNAs. Following this manual check, we tested the presence of those text mined SNPs within PolymiRTS and miRNASNP v3. The absence of recent miR-SNPs from the databases guided our independent collation step.

Collation of microRNA-associated variants

Variants within microRNA precursor and mature regions as well as those in ± 5 -kb flanking regions were collated. To achieve this, genomic coordinates of microRNA precursor and mature sequences were downloaded from miRBase v22 [63] (<https://www.mirbase.org/ftp/CURRENT/genomes/hsa.gff3>) and intersected with genomic coordinates from the full dbSNP v151 [23] catalogue using BEDTools [64].

Primary transcripts of intergenic microRNAs are not well characterised. However, several studies have shown that flanking regions between ~ 1 kb and 10 kb are likely to contain transcription start sites, CpG islands, expressed sequence tag (EST)- and transcription factor (TF)-binding sites [21, 60, 65]. By extending the microRNA precursor coordinates by ± 5 kb, we aimed to incorporate microRNA regulatory features that might be influencing microRNA expression. We extracted

sequences marked as "microRNA_primary_transcript" from the miRBase v22 gff file. These represent precursor sequences. These coordinates of these transcripts were extended by 5 kb in both directions using the BEDTools suite.

microRNA SNPs tested in summary statistics

We intersected the collated microRNA and ± 5 -kb flanking SNPs with the lifted over summary statistics. Bonferroni correction was applied on microRNA SNPs found among the summary statistics. The p value thresholds were adjusted as follows: microRNA SNPs (0.05/314) and ± 5 -kb flanking SNPs (0.05/36,841).

microRNA-associated SNPs among susceptibility SNPs

In total, 200 non-MHC autosomal SNPs were significantly associated with MS in the most recent meta-analysis [6]. Those susceptibility SNPs can be obtained from Additional file tables (Additional file 1: Table S7) of that paper. We intersected the genomic coordinates of our collated microRNA, ± 5 -kb flanking and 3'UTR target SNPs with these susceptibility SNPs. Nominally significant SNPs which did not meet the genome-wide threshold were extracted from the IMSGC [6] Additional files (Additional file 1: Table S14). These were merged with the susceptibility SNPs to create a dataset of independent SNPs. The file names for all datasets extracted from the IMSGC study are listed in Additional file 1: Table S11.

microRNA-associated SNPs in LD with susceptibility SNPs

We aimed to capture entire susceptibility loci by mapping variants in linkage disequilibrium with the susceptibility SNPs. For this step, both the effect SNPs and discovery SNPs provided in Additional file 1: Table S7 of the IMSGC analysis [6] were used as susceptibility SNPs. We obtained all variants in LD with these susceptibility SNPs through Ensembl's perl API, specifically using 1000 genomes EUR subset as the reference population. LD information was available for 174 of the 201 non-MHC susceptibility SNPs. This step was carried out for both sets of microRNA variants and the 3'UTR variants within the summary statistics.

microRNA-target gain/loss analysis

TargetScan 7.0 prediction algorithm was used locally to analyse 3'UTR binding changes in variant vs reference microRNA seed sequence. The SNP position seed sequence was located within microRNA reference FASTA sequences using SeqKit [66], which we also used to swap the reference and alternative alleles. These seed sequences were then replaced within the TargetScan miR_Family_Info.txt, while the 3'UTR file was retained. The transcripts were mapped to gene names, and

differences between the predictions for both microRNA sequences were analysed in R.

microRNA variant effect prediction

ADmiRE and miRVaS were used to predict the location and effects of microRNA variants, respectively.

Oak et al. [28] provide microRNA annotation tab files in the ADmiRE repository. These were formatted into BED files and lifted over to hg38. The BED files were intersected with vcf files of microRNA variants of interest. This procedure was implemented by Tyc and colleagues [67]. miRVaS [27] runs predictions within 2000 nucleotides of microRNA coordinates using underlying tools VARNA and RNAfold [68, 69]. miRVaS is available online, or in local Windows or Linux packages. SNP coordinates were input into miRVaS using the required format, and predictions were run based on the hg38 reference file and miRBase v21.

Collation of 3'UTR target-binding variants

TargetScan variants

We intersected TargetScan v7.0 [30] bedfiles containing genomic coordinates of all predicted sites with the UTR genome coordinates available on TargetScan. Coordinates in the former set of files were lifted over from hg19 to hg38 prior to this intersection. This intersection resulted in a collection of TargetScan-predicted binding sites within 3'UTRs. The binding site coordinates in the resulting bedfile were intersected with dbSNP 151 variants (Fig. 3A) for a final dataset of 3'UTR SNPs within binding sites predicted by TargetScan v7.0. All our intersection steps were carried out using combinations of VCFtools, BEDtools and SAMtools [70–72].

RNA22 variants

There were over 83 million predicted binding sites available from RNA22 [29] v2.0. We chose the minimum prediction p values for each microRNA–target pair predicted to interact at 3'UTRs, leading to ~14 million pairs (p value < 0.0314). Next, a custom R script was used to convert the cDNA coordinates to genomic coordinates. These were intersected with dbSNP v151 to get 1,223,207 SNPs (Fig. 3A). Additional file 2: Fig. S6 shows the overlap between targets predicted by TargetScan and this RNA “best probability” subset.

The dataset containing the union of binding site SNPs from TargetScan and RNA22 was used to test the presence of 3'UTR SNPs among the summary statistics.

3'UTR susceptibility SNPs

After intersecting the coordinates of the collated SNPs within 3'UTR binding sites with those of the susceptibility SNPs from the IMSGC, we identified 5 3'UTR binding

sites among them. Three of the transcripts relevant to the predicted microRNA-binding sites had been archived by Ensembl. Therefore, those SNPs could no longer be annotated on those transcripts. Joint p values and ORs for the two [2] candidate SNPs were obtained from Additional files (Additional file 1: Table S7) of the IMSGC 2019 meta-analysis.

Identification of independent 3'UTR SNPs

The IMSGC identified SNPs among their prioritised effects which were independent of the lead SNPs in those regions, but did not reach genome-wide significance, and were not replicated or whose joint p values were greater than the discovery p values. These SNPs are in Additional file 1: Tables S6 and S14 of the IMSGC's paper (Additional file 1: Table S11).

They identified 201 genome-wide (GW) independent effect SNPs, 117 strongly suggestive effect SNPs and 299 weakly suggestive effects. We collated a list of the weakly and strongly suggestive SNPs from these tables.

To identify independent SNPs separately from IMSGC's process, summary statistics were input into FUMA's [31] online platform (<https://fuma.ctglab.nl/>). FUMA [31] uses PLINK's [73] clumping procedures to highlight independent SNPs and lead SNPs. The intersection between both sets of independent SNPs was used for the functional prioritisation pipeline.

Among the 19 independent SNPs identified, the transcripts proposed to contain 7 3'UTR SNPs had been archived, and those SNPs had been reclassified as intronic SNPs, and the relevant 3'UTR sequences were unavailable for 2 (Additional file 1: Table S7). In addition, 2 independent SNPs had been assessed in the susceptibility SNP analysis, leaving 11 for microRNA gain/loss analysis in the next step.

microRNA–target functional pipeline

A number of groups [33–35] have proposed criteria to validate microRNA–target interactions. We have summarised these as (1) demonstration of co-expression, (2) direct interaction between miRNA and region on target, (3) gain and loss experiments to show target protein interaction and (4) predicted changes have biological functions. We have adapted these to suit our bioinformatics approach. By using in silico microRNA gain/loss, we will assess the direct interaction condition (condition 2). We will also use publically available eQTL data to meet condition 4 (changed biological functions) and are using microRNA expression data in combination with the eQTL data to test condition 1. In short, relevant 3'UTR SNPs change miRNA–target interactions at the 3'UTR binding site, act as eQTLs for the targets in MS relevant tissues (e.g. PBMCs, lymphocytes) and have the

lost/gained microRNAs expressed in the same MS relevant tissues. We are limited by study design and will not be doing the protein-level gain and loss experiments (condition 3). These criteria are highlighted in Additional file 2: Fig. S3.

We used the FiveX browser of eQTL catalogue [36] to identify the tissues in which our 3'UTR SNPs were acting as eQTLs for the predicted targets. We also checked our candidate SNPs within a more specific MS eQTL dataset which was provided alongside the MS GWAS [6] (Additional file 1: Table S11). We also identified the probability of those SNPs lie in regulatory regions within the genome through the probability score (best probability) and the type of regulatory site (RDB Rank) from RegulomeDB v.2.03 [74]. We also used the human.mirna.cellontology dataset from FANTOM5 [38] to check which cells our miRNAs were enriched/depleted in. In addition, checked the basal expression on the webtool Zenbu miRNA atlas (comparing microRNA expression across 0.5 low/10 medium/1000 highTPM) (Additional file 2: Fig. S3).

microRNA gain/loss analysis

We used TargetScan 7.0 prediction algorithm locally to analyse microRNA-binding changes in variant vs reference 3'UTR sequences. The TargetScan miR_Family_Info.txt file was retained, while the reference and alternative 3'UTR sequences were formatted using SeqKit [66] to match and replace the 3'UTR file. We compared the predicted microRNA families compared between output files from the alternative and reference sequences in an R script.

Abbreviations

| | |
|-------|---|
| 3'UTR | 3' Untranslated region |
| CNS | Central nervous system |
| CSF | Cerebrospinal fluid |
| EAE | Experimental autoimmune encephalomyelitis |
| EBV | Epstein–Barr virus |
| GWAS | Genome-wide association studies |
| LD | Linkage disequilibrium |
| MHC | Major histocompatibility complex |
| MS | Multiple sclerosis |
| miRNA | MicroRNA |

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40246-023-00478-4>.

Additional file 1: Table S1. miRNA SNPs in summary statistics. **Table S2.** RNAfold rs1414273 effects. **Table S3.** MIR6891 validated targets lost. **Table S4.** MIR6891-predicted targets lost. **Table S5.** Candidate SNPs in flanking regions of miRNA precursors. **Table S6.** miRNA gain/loss rs6742. **Table S7.** Independent SNPs. **Table S8.** miRNA gain/loss independent SNPs. **Table S9.** Functional prioritisation. **Table S10.** miRNA-associated SNP candidates. **Table S11.** List of IMSCG data files referenced.

Additional file 2. Supplementary Figures 1–6.

Acknowledgements

Not applicable

Author contributions

IF was involved in analysis, writing the original draft, writing—reviewing and editing, and visualisation, CMC was responsible for conceptualisation, supervision and writing—reviewing and editing. SF was responsible for supervision and writing—reviewing and editing. All authors read and approved the final manuscript.

Funding

This project was funded by Science Foundation Ireland under the SFI Centre for Research Training in Genomics Data Science grant 18/CRT/6214.

Availability of data and materials

MS GWAS meta-analysis summary statistics are available at the IMSCG website: <http://imsgc.net/>. The data, which support the conclusions of this study, are included in this published article and its Additional files. Other datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹School of Pharmacy and Biomolecular Sciences, Royal College of Surgeons in Ireland, Dublin, Ireland. ²SFI Centre for Research Training in Genomics Data Sciences, University of Galway, H91 TK33 Galway, Ireland. ³FutureNeuro SFI Research Centre, Royal College of Surgeons in Ireland, Dublin, Ireland. ⁴Genomic Oncology Research Group, Department of Physiology & Medical Physics, Royal College of Surgeons in Ireland, Dublin, Ireland.

Received: 14 February 2023 Accepted: 16 March 2023

Published online: 30 March 2023

References

- Oksenberg JR, Baranzini SE, Sawcer S, Hauser SL. The genetics of multiple sclerosis: SNPs to pathways to pathogenesis. *Nat Rev Genet.* 2008;9(7):516–26.
- Brownlee WJ, Hardy TA, Fazekas F, Miller DH. Diagnosis of multiple sclerosis: progress and challenges. *Lancet.* 2017;389(10076):1336–46.
- Hollenbach JA, Oksenberg JR. The immunogenetics of multiple sclerosis: A comprehensive review. *J Autoimmun.* 2015;64:13–25.
- Bjornevik K, Cortese M, Healy BC, Kuhle J, Mina MJ, Leng Y, et al. Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science.* 2022;301(January):296–301.
- Cotsapas C, Mitrovic M. Genome-wide association studies of multiple sclerosis. *Clin Transl Immunol.* 2018;7(6):1–9.
- Patsopoulos NA, Baranzini SE, Santaniello A, Shoostari P, Cotsapas C, Wong G, et al. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science.* 2019;365:6460.
- Mitrović M, Patsopoulos NA, Beecham AH, Dankowski T, Goris A, Dubois B, et al. Low-frequency and rare-coding variation contributes to multiple sclerosis risk. *Cell.* 2018;175(6):1679–1687.e7.
- Kehl T, Backes C, Kern F, Fehlmann T, Ludwig N, Meese E, et al. About miRNAs, miRNA seeds, target genes and target pathways. *Oncotarget.* 2017;8(63):107167–75.
- O'Brien J, Hayder H, Zayed Y, Peng C. Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Front Endocrinol.* 2018. <https://doi.org/10.3389/fendo.2018.00402>.

10. Jużwik CA, Drake SS, Zhang Y, Paradis-Isler N, Sylvester A, Amar-Zifkin A, Douglas C, Morquette B, Moore CS, Fournier AE. microRNA dysregulation in neurodegenerative diseases: a systematic review. *Progress Neurobiol.* 2019;182:101664. <https://doi.org/10.1016/j.pneurobio.2019.101664>.
11. Teuber-Hanselmann S, Meinel E, Junker A. MicroRNAs in gray and white matter multiple sclerosis lesions: impact on pathophysiology. *J Pathol.* 2020;250(5):496–509.
12. Rhead B, Shao X, Graves JS, Chitnis T, Waldman AT, Lotze T, et al. miRNA contributions to pediatric-onset multiple sclerosis inferred from GWAS. *Ann Clin Transl Neurol.* 2019;6(6):1053–61.
13. Afrasiabi A, Fewings NL, Schibeci SD, Keane JT, Booth DR, Parnell GP, et al. The interaction of human and Epstein–Barr virus miRNAs with multiple sclerosis risk loci. *Int J Mol Sci.* 2021;22(6):1–15.
14. Hecker M, Boxberger N, Illner N, Fitzner B, Schröder I, Winkelmann A, Dudesek A, Meister S, Koczan D, Lorenz P, Thiesen H-J, Zettl UK. 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. <https://doi.org/10.1371/journal.pgen.1007961>.
15. Dehghanzad R, Panahi Moghadam S, Shirvani FZ. Prediction of single-nucleotide polymorphisms within microRNAs binding sites of neuronal genes related to multiple sclerosis: a preliminary study. *Adv Biomed Res.* 2021;10:8.
16. Landi D, Gemignani F, Naccarati A, Pardini B, Vodicka P, Vodickova L, et al. Polymorphisms within microRNA-binding sites and risk of sporadic colorectal cancer. *Carcinogenesis.* 2008;29(3):579–84.
17. Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet.* 2011;12(12):861–74.
18. Ghanbari M, Franco OH, de Looper HWJ, Hofman A, Erkeland SJ, Dehghan A. Genetic variations in MicroRNA-binding sites affect MicroRNA-mediated regulation of several genes associated with cardio-metabolic phenotypes. *Circ Cardiovasc Genet.* 2015;8(3):473–86.
19. Ghanbari M, Iglesias AI, Springelkamp H, van Duijn CM, Ikram MA, Dehghan A, et al. A genome-wide scan for microRNA-related genetic variants associated with primary open-angle glaucoma. *Investig Ophthalmol Vis Sci.* 2017;58(12):5368–77.
20. Gholami M, Zoughi M, Larijani B, Amoli MM, Bastami M. An in silico approach to identify and prioritize miRNAs target sites polymorphisms in colorectal cancer and obesity. *Cancer Med.* 2020;9(24):9511–28. <https://doi.org/10.1002/cam4.3546>.
21. Fang L, Sørensen P, Sahana G, Panitz F, Su G, Zhang S, et al. MicroRNA-guided prioritization of genome-wide association signals reveals the importance of microRNA-target gene networks for complex traits in cattle. *Sci Rep.* 2018;8(1):1–14.
22. Ghanbari M, Ikram MA, De Looper HWJ, Hofman A, Erkeland SJ, Franco OH, et al. Genome-wide identification of microRNA-related variants associated with risk of Alzheimer's disease. *Sci Rep.* 2016;6(January):1–9.
23. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001;29(1):308–11.
24. Kozomara A, Griffiths-Jones S. MiRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 2014;42(D1):D68–73.
25. Liu C-J, Xin F, Xia M, Zhang Q, Zhifeng G, Guo A-Y. miRNASNP-v3: a comprehensive database for SNPs and disease-related variations in miRNAs and miRNA targets. *Nucl Acids Res.* 2021;49(D1):D1276–81. <https://doi.org/10.1093/nar/gkaa783>.
26. Bhattacharya A, Ziebarth JD, Cui Y. PolymIRTS database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res.* 2014;42(D1):D86–91.
27. Cammaerts S, Strazisar M, Dierckx J, Del Favero J, De Rijck P. miRVaS: a tool to predict the impact of genetic variants on miRNAs. *Nucleic Acids Res.* 2016;44(3):e23–e23.
28. Oak N, Ghosh R, Huang K, et al. Framework for microRNA variant annotation and prioritization using human population and disease datasets. *Hum Mutat.* 2019;40(1):73–89.
29. Miranda KC, Huynh T, Tay Y, Ang Y-S, Tam W-L, Thomson AM, et al. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell.* 2006;126(6):1203–17.
30. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife.* 2015;4:e05005.
31. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun.* 2017;8(1):1–10.
32. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559–75. <https://doi.org/10.1086/519795>.
33. Riolo G, Cantara S, Marzocchi C, Ricci C. miRNA targets: From prediction tools to experimental validation. *Methods Protoc.* 2021;4(1):1–20.
34. Kuhn DE, Martin MM, Feldman DS, Terry AV, Nuovo GJ, Elton TS. Experimental validation of miRNA targets. *Methods.* 2008;44(1):47–54.
35. Elton TS, Yalowich JC. Experimental procedures to identify and validate specific mRNA targets of miRNAs. *EXCLI J.* 2015;14:758–90.
36. Kerimov N, Hayhurst JD, Peikova K, Manning JR, Walter P, Kolberg L, et al. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat Genet.* 2021;53(9):1290–9.
37. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated encyclopedia of DNA elements in the human genome. *Nature.* 2012;489(7414):57–74.
38. Forrest ARR, Kawaji H, Rehli M, Kenneth Baillie J, de Hoon MJL, Haberle V, et al. A promoter-level mammalian expression atlas. *Nature.* 2014;507(7493):462–70.
39. Quach H, Rotival M, Pothlichet J, et al. Genetic adaptation and neandertal admixture shaped the immune system of human populations. *Cell.* 2016;167(3):643–656.e17. <https://doi.org/10.1016/j.cell.2016.09.024>.
40. Chen Y, Knight ZA. Making sense of the sensory regulation of hunger neurons. *BioEssays.* 2016;38(4):316–24. <https://doi.org/10.1002/bies.20150167>.
41. Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martín D, et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell.* 2016;167(5):1398–1414.e24.
42. Kwong A, Boughton AP, Wang M, VandeHaar P, Boehnke M, Abecasis G, et al. FIVEx: an interactive eQTL browser across public datasets. *Bioinformatics.* 2022;38(2):559–61.
43. Schmiedel BJ, Singh D, Madrigal A, Valdovino-Gonzalez AG, White BM, Zapardiel-Gonzalo J, et al. Impact of genetic polymorphisms on human immune cell gene expression. *Cell.* 2018;175(6):1701–1715.e16.
44. Jaffe AE, Straub RE, Shin JH, Tao R, Gao Y, Collado-Torres L, et al. Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat Neurosci.* 2018;21(8):1117–25.
45. Lappalainen T, et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature.* 2013;501(7468):506.
46. Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A, et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife.* 2013;2:e00523.
47. Lepik K, Annilo T, Kukuškina V, Kisand K, Kutalik Z, Peterson P, Peterson H. C-reactive protein upregulates the whole blood expression of CD59 - an integrative analysis. *PLOS Comput Biol.* 2017;13(9):e1005766. <https://doi.org/10.1371/journal.pcbi.1005766>.
48. Hecker M, Fitzner B, Putscher E, Schwartz M, Winkelmann A, Meister S, Dudesek A, Koczan D, Lorenz P, Boxberger N, Zettl UK. Implication of genetic variants in primary microRNA processing sites in the risk of multiple sclerosis. *EBioMedicine.* 2022;80:104052. <https://doi.org/10.1016/j.ebiom.2022.104052>.
49. Sakaue S, Hirata J, Maeda Y, Kawakami E, Nii T, Kishikawa T, et al. Integration of genetics and miRNA-target gene network identified disease biology implicated in tissue specificity. *Nucleic Acids Res.* 2018;46(22):11898–909.
50. Jacinta-Fernandes A, Xavier JM, Magno R, Lage JG, Maia A-T. Allele-specific miRNA-binding analysis identifies candidate target genes for breast cancer risk. *npj Genom Med.* 2020. <https://doi.org/10.1038/s41525-019-0112-9>.
51. Shieh M, Chitnis N, Clark P, Johnson FB, Kamoun M, Monos D. Computational assessment of miRNA binding to low and high expression HLA-DPB1 allelic sequences. *Hum Immunol.* 2019;80(11):53–61.
52. Hauberg ME, Holm-Nielsen MH, Mattheisen M, Askou AL, Grove J, Børglum AD, et al. Schizophrenia risk variants affecting microRNA function and site-specific regulation of NT5C2 by miR-206. *Eur Neuropsychopharmacol.* 2016;26(9):1522–6.
53. Isoke N, Madireddy L, Khankhanian P, Matsushita T, Caillier SJ, Moré JM, et al. An ImmunoChip study of multiple sclerosis risk in African Americans. *Brain.* 2015;138(6):1518–30.

54. Healy BC, Liguori M, Tran D, Chitnis T, Glanz B, Wolfish C, Gauthier S, Buckle G, Houtchens M, Stazzone L, Khoury S, Hartzmann R, Fernandez-Vina M, Hafler DA, Weiner HL, Guttman CRG, De Jager PL. HLA B*44: Protective effects in MS susceptibility and MRI outcome measures. *Neurology*. 2010;75(7):634–40. <https://doi.org/10.1212/WNL.0b013e3181ed9c9c>.
55. Moutsianas L, Jostins L, Beecham AH, Dilthey AT, Xifara DK, Ban M, et al. Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nat Genet*. 2015;47(10):1107–13.
56. Zhang P, Sun J, Liang C, Gu B, Xu Y, Lu H, et al. lncRNA IGHcy1 Acts as a ceRNA to Regulate Macrophage Inflammation via the miR-6891-3p/TLR4 Axis in Osteoarthritis. *Mediators Inflamm*. 2020;17(2020):9743037.
57. Lang F, Singh Y, Salker MS, Ma K, Pandya AA, Lang PA, et al. Glucose transport in lymphocytes. *Pflügers Arch - Eur J Physiol*. 2020;472(9):1401–6.
58. Lou G, Palikaras K, Lautrup S, Scheibye-Knudsen M, Tavernarakis N, Fang EF. Mitophagy and neuroprotection. *Trends Mol Med*. 2020;26(1):8–20.
59. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res*. 2002;12(6):996–1006.
60. Saini HK, Griffiths-Jones S, Enright AJ. Genomic analysis of human microRNA transcripts. *Proc Natl Acad Sci U S A*. 2007;104(45):17719–24.
61. Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, et al. Ensembl 2021. *Nucleic Acids Res*. 2021;49(D1):D884–91.
62. Boughton AP, Welch RP, Flickinger M, VandeHaar P, Taliun D, Abecasis GR, Boehnke M. LocusZoom.js: interactive and embeddable visualization of genetic association study results. *Bioinformatics*. 2021;37(18):3017–8. <https://doi.org/10.1093/bioinformatics/btab186>.
63. Kozomara A, Birgaoanu M, Griffiths-Jones S. MiRBase: From microRNA sequences to function. *Nucleic Acids Res*. 2019;47(D1):D155–62.
64. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.
65. Gong J, Tong Y, Zhang HM, Wang K, Hu T, Shan G, et al. Genome-wide identification of SNPs in MicroRNA genes and the SNP effects on MicroRNA target binding and biogenesis. *Hum Mutat*. 2012;33(1):254–63.
66. Shen W, Le S, Li Y, Hu F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE*. 2016;11(10): e0163962.
67. Tyc KM, Wong A, Scott RT, Tao X, Schindler K, Xing J. Analysis of DNA variants in miRNAs and miRNA 3'UTR binding sites in female infertility patients. *Lab Invest*. 2021;101(4):503–12.
68. Lorenz R, et al. ViennaRNA package 2.0. *Algorithms Mol Biol*. 2011;6(1):26.
69. Darty K, Denise A, Ponty Y. VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics*. 2009;25(15):1974–5.
70. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCftools. *Gigascience*. 2021;10:2.
71. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
72. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–8.
73. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015. <https://doi.org/10.1186/s13742-015-0047-8>.
74. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22(9):1790–7.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

