Integrating the Ribonucleic Acid Sequencing Data From Various Studies for Exploring the Multiple Sclerosis-Related Long Noncoding Ribonucleic Acids and Their Functions

Zhijie Han[1,2], Jiao Hua[3], Weiwei Xue[2] and Feng Zhu[1,2]*

[1] College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, China, [2] School of Pharmaceutical Sciences, Chongqing University, Chongqing, China, [3] School of Mathematics, Harbin Institute of Technology, Harbin, China

Multiple sclerosis (MS) is a chronic fatal central nervous system (CNS) disease involving in complex immunity dysfunction. Recently, long noncoding RNAs (lncRNAs) were discovered as the important regulatory factors for the pathogenesis of MS. However, these findings often cannot be repeated and confirmed by the subsequent studies. We considered that the small-scale samples or the heterogeneity among various tissues may result in the divergence of the results. Currently, RNA-seq has become a powerful approach to quantify the abundances of lncRNA transcripts. Therefore, we comprehensively collected the MS-related RNA-seq data from a variety of previous studies, and integrated these data using an expression-based meta-analysis to identify the differentially expressed lncRNA between MS patients and controls in whole samples and subgroups. Then, we performed the Jensen-Shannon (JS) divergence and cluster analysis to explore the heterogeneity and expression specificity among various tissues. Finally, we investigated the potential function of identified lncRNAs for MS using weighted gene co-expression network analysis (WGCNA) and gene set enrichment analysis (GSEA), and 5,420 MS-related lncRNAs specifically expressed in the brain tissue were identified. The subgroup analysis found a small heterogeneity of the lncRNA expression profiles between brain and blood tissues. The results of WGCNA and GSEA showed that a potential important function of lncRNAs in MS may be involved in the regulation of ribonucleoproteins and tumor necrosis factor cytokines receptors. In summary, this study provided a strategy to explore disease-related lncRNAs on genome-wide scale, and our findings will be benefit to improve the understanding of MS pathogenesis.

Keywords: ribonucleic acid sequencing, multiple sclerosis, long non-coding ribonucleic acids, meta-analysis, function analysis

INTRODUCTION

Multiple sclerosis (MS) is a chronic fatal neurodegenerative disease involving in complex immunity [central nervous system (CNS)] (Sospedra and Martin, 2005; Frohman et al., 2006; Li et al., 2018). Based on the 2014 statistics of the Atlas of MS investigation, the estimated number of the people afflicted with the MS worldwide has reached approximately 2.3 million (Browne et al., 2014). Although much remains unknown about the molecular etiology of MS, more and more studies showed that the dysregulation of transcriptional processes could potentially contribute to the pathogenesis of MS (Li et al., 2017; Selmaj et al., 2017; Angerer et al., 2018; Cheng et al., 2018; Han et al., 2018b; Zhang et al., 2019).

Recently, long noncoding RNA (lncRNA), one of the non-protein-coding genes whose transcripts are longer than 200 nucleotides, has been discovered as the important regulatory factor of immune system and pathogenesis of CNS disorders including MS (Gomez et al., 2013; Ng et al., 2013; Dong et al., 2015; Cheng et al., 2016; Santoro et al., 2016; Zhang et al., 2016; Chen et al., 2017; Eftekharian et al., 2017; He et al., 2017; Cheng et al., 2018; Yin et al., 2019). However, for MS, these results often cannot be repeated and confirmed by subsequent study. For example, multiple variants of the lncRNA antisense non-coding RNA in the INK4 locus (*ANRIL*) are found significantly associated with the risk of MS through the haplotype analysis of blood samples (Rezazadeh et al., 2018). But following study reveals that the function of *ANRIL* does not contribute the pathogenesis of MS in blood, cortex, and cerebellum tissues (Pahlevan Kakhki et al., 2018). Study showed a significant upregulation of lncRNA *MALAT1* in MS blood tissues (Cardamone et al., 2019), while the expression of *MALAT1* was found markedly decreased in MS brain by the subsequent study (Masoumi et al., 2019). Moreover, another study found that *MALAT1* is not significantly differentially expressed between MS patients and controls (Gharesouran et al., 2019). We considered that the small-scale samples or the heterogeneity among various tissues may result in the divergence of the results.

Currently, specifically for lncRNAs, using RNA-seq data to quantify abundance of the transcripts has become very powerful approach compared with the traditional ones (e.g., gene microarray) (Wang et al., 2009). Particularly, almost all of the expression of the known lncRNA transcripts can be measured using RNA-seq data, but this proportion is just approximately 0.1 to 10.6% by the method of probe re-annotation using various types of microarrays (Du et al., 2013; Fang et al., 2018; Yang et al., 2019). Moreover, lncRNA abundance quantification using RNA-seq data also shows higher accuracy based on its deep read coverage, while the re-annotation approach only requires the sequence match of 1 to 4 probes when quantifies lncRNA abundance (Du et al., 2013; Gellert et al., 2013; Li et al., 2019). A previous study reported that by paying attention to some aspect of library and sequencing process [i.e., poly-A tail selection, paired-end sequencing, and sequencing of double-stranded complementary DNA (cDNA)], the lncRNAs are more easily and more accurately identified through RNA-seq (Ilott and Ponting, 2013).

In this study, we thus selected all MS-related RNA-seq data in a variety of studies by searching three authoritative public databases: GEO DataSets (Barrett et al., 2013), EBI-EMBL ArrayExpress (Athar et al., 2019), and DDBJ Sequence Read Archive (Ogasawara et al., 2013) using the keyword "multiple sclerosis." Then, we used these RNA-seq data to perform expression quantification of the lncRNA in each of the selected studies. Next, we integrated the lncRNA expression results of all selected studies by an expression-based meta-analysis to identify the significantly differentially expressed lncRNAs between MS patients and controls. Further, we explored their heterogeneity and expression specificity among various tissues. After that, the weighted gene co-expression network analysis (WGCNA) was performed using the expression data of lncRNAs and protein-coding genes to identify the significant modules for MS. The expression of the protein-coding genes was calculated using the same approach on lncRNA. Finally, we conducted gene set enrichment analysis (GSEA) on the co-expressed protein-coding genes in each significant module to infer the function of the differentially expressed lncRNAs potentially contributing to the pathogenesis of MS.
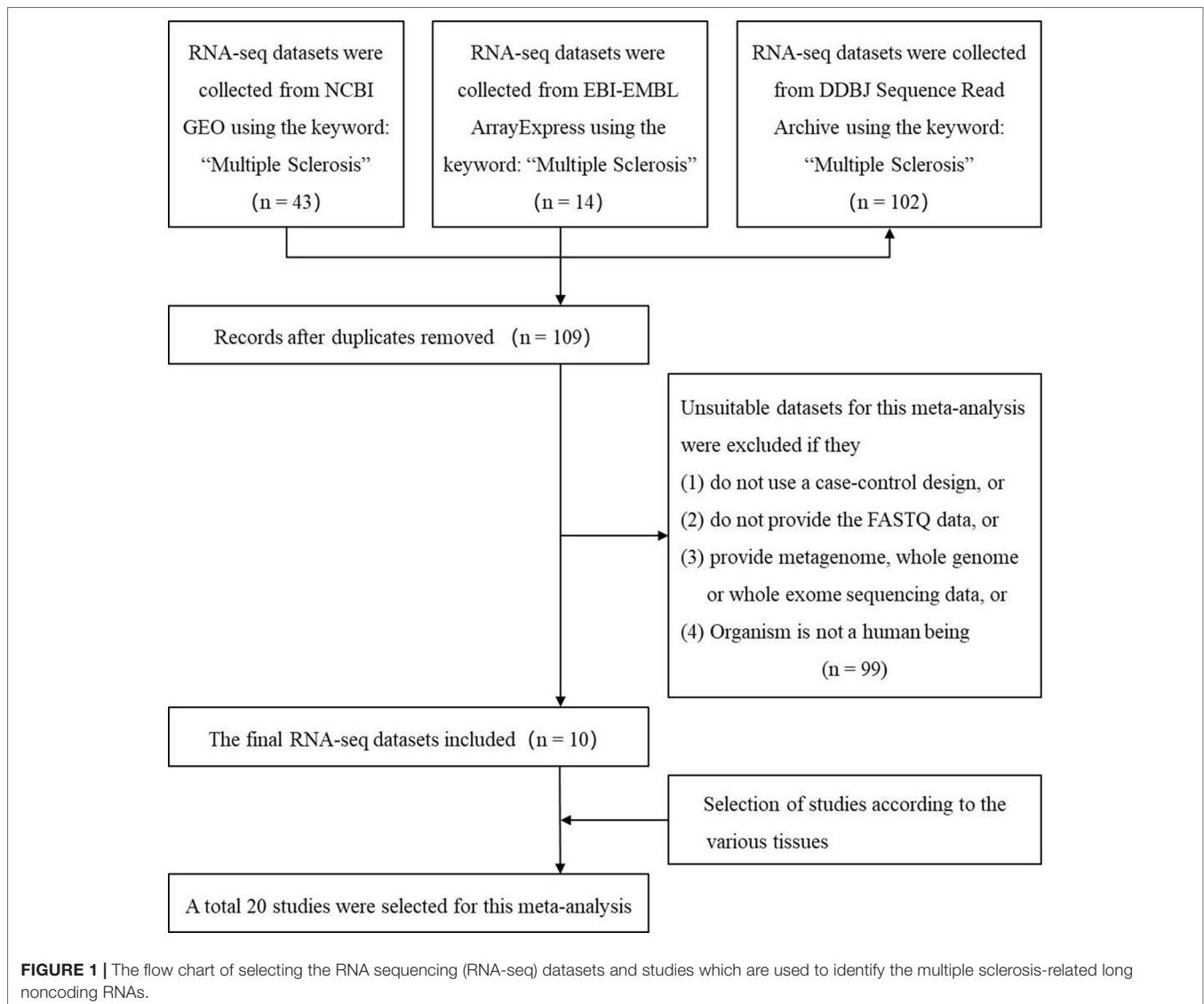
MATERIALS AND METHODS

Selection of the Multiple Sclerosis-Related Ribonucleic Acid Sequencing Datasets and Studies

We used the keyword "multiple sclerosis" to search all the possible MS-related RNA-seq datasets in three authoritative databases: GEO DataSets (Barrett et al., 2013), EBI-EMBL ArrayExpress (Athar et al., 2019), and DDBJ Sequence Read Archive (Ogasawara et al., 2013). The search was performed before the last update of the databases on May 16 2019. Then, we selected the suitable datasets using four criteria: 1) the organism in the dataset is the human being; 2) the study in the dataset is designed using the case-control method; 3) the dataset has provided the FASTQ data; (4) the FASTQ data in the dataset is not generated by metagenome, whole genome, or whole exome sequencing. Finally, the studies from these datasets based on various tissues were selected. **Figure 1** showed the workflow.

Quantification of Long Noncoding Ribonucleic Acid Sequencing Abundance Using Ribonucleic Acid Sequencing Sequencing Data

We first downloaded the sequence data of these studies by *Prefetch* and converted them into FASTQ files using *fastq-dump* tool of the SRA Toolkit software (Leinonen et al., 2011). Next, we downloaded the reference sequences of lncRNA and protein-coding transcripts in FASTA format from NONCODE (version 5) (Fang et al., 2018) and Ensembl (release 91) (Aken et al., 2017), respectively, and further merged the two FASTA format files. Particularly, NONCODE is one of the most complete and well-annotated databases of the noncoding RNAs, and we obtained a total of 172,216 transcript sequences of 96,308 human

FIGURE 1 | The flow chart of selecting the RNA sequencing (RNA-seq) datasets and studies which are used to identify the multiple sclerosis-related long noncoding RNAs.

lncRNA genes from it. Ensembl aggregated the cDNA data from National Center for Biotechnology Information (Sayers et al., 2019), UniProt (UniProt, 2015), Genome Reference Consortium (Church et al., 2011), and UCSC Genome Browser (Kent et al., 2002) databases. After removing the pseudogenes, we obtained a total of 160,040 transcript sequences of 22,810 human protein-coding genes from it. Then, we performed the quantification of the lncRNA and protein-coding transcripts simultaneously by mapping the RNA-seq reads of each study to the merged reference sequence (pseudoalignment) and calculating the count values using *Kallisto* software (Bray et al., 2016). *Kallisto* is a fast and highly accurate quantification tool for transcript abundance through k-mer lookup technique. Here, the merged reference sequences have been processed into a transcriptome index to conduct the pseudoalignment which has the same effect as the reads alignment to a given reference genome in the traditional transcript-level RNA-seq processing but can substantially reduce calculation time. For the paired-end sequencing samples, the arguments were set to defaults, i.e., the number of bootstrap samples (-b) equals 0 and the number of threads (-t) equals 1. For the single-end sequencing samples, besides these default parameter settings, we set the estimated average fragment length (-l) and the standard deviation of fragment length (-s) to 200 and 20, respectively, according to Kallisto's recommended parameters. Finally, based on the annotation file "Transcript2Gene," we integrated transcript-level count values of lncRNAs to calculate their corresponding gene-level count values using the R package "tximport" (Soneson et al., 2015).

Heterogeneity Test and Meta-Analysis

To identify the significantly differentially expressed lncRNAs between MS patients and controls, we calculated and integrated the results of each study by a meta-analysis. These analyses were conducted using R package "MetaOmics," which is a comprehensive analytical pipeline to meta-analyze multiple

transcriptomic studies (Ma et al., 2019). This meta-analysis includes a normalization process same as the edgeR's strategy and a "AW-Fisher" method to integrate data (Bullard et al., 2010; Robinson et al., 2010; Ma et al., 2019). First, we calculated the two parameters, I^2 and P value, to measure the lcnRNA expression heterogeneity by the Cochran's Q Statistics, which is based on a chi-square test with $k − 1$ degrees of freedom (k equals to the number of studies used for the meta-analysis). According to the previous studies, the heterogeneity was considered as statistically significant when $I^2 > 50\%$ and $P < 0.01$ (Han et al., 2015; Li et al., 2016; Liu et al., 2017; Han et al., 2018a; Xue et al., 2018). Then, the meta-analysis was performed for each of these lncRNAs based on their count values. Particularly, the random effect model (REM) and fixed effect model (FEM) were used, respectively, for the lncRNAs with a significant heterogeneity or not. Using the REM in meta-analysis can reduce bias of the results (Kim et al., 2015; Szajewska and Kolodziej, 2015). We calculated standardized mean difference (SMD) with its 95% confidence interval (CI) to identify the differentially expressed lncRNA between the MS patients and controls (95% CI of SMD does not include zero, FDR adjusted $P < 0.05$). The SMD is given by the mean difference between case and control divided by the standard deviation and applies to meta-analysis when the outcome is continuous variable (e.g., expression level). Moreover, since all these samples can be split into brain and blood, we performed the meta-analysis for the two subgroups, and explored the differential expression pattern of the MS-related lncRNAs between brain and blood.

In addition, we further explored the specific target genes of the lncRNAs using LncRNA2Target v2.0 database which is authoritative source including 152,137 lncRNA-target relationships confirmed by the knockdown or overexpression analysis and binding experimental technologies, and provides web interface for searching the targets by a particular lncRNA (Cheng et al., 2019).

Tissue Specificity Analysis of the Multiple Sclerosis-Related Long Noncoding Ribonucleic Acids

We explored the tissue expression specificity of the significantly differentially expressed lncRNAs in MS, which was important aspects of neurological disease research (usually, specifically expressed in CNS system) (Fatica and Bozzoni, 2014; He et al., 2017; Tang et al., 2019b). For this purpose, lncRNA expression data were first downloaded from the NONCODE, which were involved in primary human tissue/cell line (e.g., brain, heart, breast, lung, liver, foreskin, lung, lymph node, colon, skeletal muscle, leukocyte, HeLa cells, and fibroblasts, etc.). Then, we extracted the expression data of various tissues by the corresponding differentially expressed lncRNAs in brain, blood, and whole sample, respectively, and stored them in three independent sets. Further, based on these data, we used the Jensen-Shannon (JS) divergence, an entropy-based approach, to calculate a tissue specificity score of the differentially expressed lncRNAs according to previous study (Cabili et al., 2011). Briefly, the lncRNA expression vectors were converted to abundance density, and the distance between two tissue expression patterns was defined as the square root of JS divergence. The tissue

specificity of a lncRNA expression pattern was measured through the distance between expression patterns across various tissues and predefined extreme pattern in which the lncRNA is uniquely expressed in one tissue (1 minus the distance). Thus, the metric of tissue specificity ranged from 0 to 1. The nearer the score to one, the stronger the tissue specificity becomes. Finally, using the same data, we performed the cluster analysis with Manhattan distance for differentially expressed lncRNAs in brain, blood and whole sample by R package "gplots."

Inferring the Functions of Multiple Sclerosis-Related Long Noncoding Ribonucleic Acids by Weighted Gene Co-Expression Network Analysis

To infer the potential biological functions of these significantly differentially expressed lncRNAs in MS, we used WGCNA approach to determine the co-expression profile of these MS-related lncRNAs and protein-coding genes, and further performed the GSEA by the co-expressed protein-coding genes. First, in the same way used for identifying MS-related lncRNAs, we quantified the abundance of the protein-coding genes and identified the significantly differentially expressed genes by a meta-analysis. Second, we constructed the co-expression network by integrating the count values of the differentially expressed lncRNAs and protein-coding genes using the R package "WGCNA" (Langfelder and Horvath, 2008). Particularly: 1) we conducted the sample clustering to check if there were any outlier samples using "hclust" function of R package "WGCNA"; 2) after quality control, we used "pickSoftThreshold" function of R package "WGCNA" to calculate the satisfactory soft threshold power β for ensuring the scale-free topology characteristics of the co-expression network; 3) based on the β value, we applied the Pearson's method to calculate an adjacency matrix which includes the weighted correlation of all gene pairs; 4) by adjacency matrix, we used the dynamic cut-tree algorithm to construct a hierarchical clustering dendrogram and identified the co-expression modules where genes have high topological overlap with each other. Finally, we assessed the significance of the modules for MS by measuring two indices. Particularly, one of the indices is correlation between module membership (i.e., intramodular connectivity) and gene significance for MS. High correlation means that the hub genes (i.e., the genes with high connectivity in a co-expression module) of the corresponding module also tend to be highly correlated with disease states (MS or healthy) (Langfelder and Horvath, 2008). The other is the average correlation of the genes in each module with disease states. This was also applied to assess association of each module with the platforms and the tissue types, respectively.

Pathway Analysis of the Multiple Sclerosis-Related Long Noncoding Ribonucleic Acids by Gene Set Enrichment Analysis

Based on the two indices of module significance, we selected the most significant modules of disease states to investigate the

lncRNA functions in MS by GSEA. We first extracted the ID numbers of the protein-coding genes co-expressed with lncRNAs in the modules. Then, we downloaded the signaling pathway data from two common databases, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). GO is a public resource of data on the gene functions in the biological process, molecular function, and cellular component (The Gene Ontology, 2017), and KEGG is comprehensive database which integrates the information of genes involved in signaling pathways, cellular processes, human diseases, etc. (Kanehisa et al., 2017). Finally, we used the co-expressed protein-coding genes and the signaling pathway data to conduct the GSEA of the most significant modules using R package "clusterProfiler" (Yu et al., 2012). The adjusted P value calculated by the multiple testing (Benjamini-Hochberg method) was set at less than 0.05 as the threshold of significance.

RESULTS AND DISCUSSION

Results of Study Selection and Long Noncoding Ribonucleic Acid Abundance Quantification

Using keyword search and quality filtering, we identified ten MS-related RNA-seq datasets including: GSE60424, GSE66573, GSE66763, GSE89843, GSE100297, GSE120411, GSE111972, GSE123496, GSE77598, and SRP132699 from three authoritative databases. We found that the library preparation and sequencing methods in most of these datasets meet one/multiple requirements for improving the

lncRNAs identification (i.e., poly-A tail selection, paired-end sequencing, and sequencing of double-stranded cDNA). Then, after the investigating the source of samples, we found that these datasets are involved in eight brain tissues (optic chiasm, corpus callosum, occipital cortex, astrocytes, frontal cortex, hippocampus, internal capsule, parietal cortex) and seven blood tissues (B cell, T cell, monocyte, platelets, neutrophils, natural killer cell, and whole blood). According to the various tissues, we selected a total of 20 studies (207 MS cases and 348 controls) for the following analysis. The detailed information of each study was shown in **Table 1**. Finally, we downloaded RNA-seq data of the samples in each study, and used them to measure lncRNA expression (count values) using *Kallisto* (Bray et al., 2016) and R package "tximport" (Soneson et al., 2015). In total, lncRNA abundance in 555 samples was quantified.

Heterogeneity Test and Meta-Analysis

Based on the count values of the 96,308 lncRNAs in 20 studies, the meta-analysis was performed to calculate SMD value with its 95% CI for each lncRNA using REM/FEM. Heterogeneity test showed that only about 2.90% lncRNAs have the significant heterogeneity ($I^2 > 50\%$ and $P < 0.01$). Therefore, the homogeneous unbiased results could be identified in >97% lncRNAs by FEM. For the remaining lncRNAs of significant heterogeneity, REM could reduce resulting bias. In total, 5,420 lncRNAs were identified significantly differentially expressed between MS cases and controls, which included 368 downregulated and 5,052 upregulated lncRNAs (shown in **Figure 2A** and **Supplementary Table S1**). For example, the **Figure 2B** exhibited the meta-analysis results of the lncRNA NONHSAG108980.1 which shows the

TABLE 1 | Summary of the 20 selected studies for the meta-analysis. NK, natural killer cell.

| Study Number	Dataset	Tissue	Year	No. of cases	No. of controls	Sequencing platform	RNA-seq library type		
							Poly-A tail select	Sequencing of double-stranded cDNA	Read type
1	GSE60424	B-cells	2014	6	4	Illumina HiScanSQ	Yes	Not described	Paired-end
2	GSE60424	Monocytes	2014	6	4	Illumina HiScanSQ	Yes	Not described	Paired-end
3	GSE60424	Neutrophils	2014	6	4	Illumina HiScanSQ	Yes	Not described	Paired-end
4	GSE60424	NK	2014	3	4	Illumina HiScanSQ	Yes	Not described	Paired-end
5	GSE60424	T-cells	2014	12	8	Illumina HiScanSQ	Yes	Not described	Paired-end
6	GSE60424	Whole blood	2014	6	4	Illumina HiScanSQ	Yes	Not described	Paired-end
7	GSE66573	Whole blood	2015	6	8	Illumina HiSeq 2500	Yes	Yes	Paired-end
8	GSE66763	T-cells	2015	10	6	Illumina HiSeq 2500	Not described	Not described	Paired-end
9	GSE77598	Monocytes	2016	5	3	Illumina HiSeq 2000	Not described	Not described	Paired-end
10	GSE89843	Platelets	2017	58	234	Illumina HiSeq 2500	Yes	Yes	Single-end
11	GSE100297	Optic chiasm	2017	5	5	Illumina HiSeq 3000	Yes	Yes	Single-end
12	GSE111972	Corpus callosum	2018	10	11	Illumina NextSeq 500	Yes	Not described	Single-end
13	GSE111972	Occipital cortex	2018	5	5	Illumina NextSeq 500	Yes	Not described	Single-end
14	GSE120411	Astrocytes	2018	24	18	Illumina HiSeq 2500	Yes	Not described	Single-end
15	SRP132699	Monocytes	2018	20	5	Illumina HiSeq 2500	Not described	Not described	Single-end
16	GSE123496	Corpus callosum	2019	5	5	Illumina HiSeq 3000	Yes	Yes	Single-end
17	GSE123496	Frontal cortex	2019	5	5	Illumina HiSeq 3000	Yes	Yes	Single-end
18	GSE123496	Hippocampus	2019	5	5	Illumina HiSeq 3000	Yes	Yes	Single-end
19	GSE123496	Internal capsule	2019	5	5	Illumina HiSeq 3000	Yes	Yes	Single-end
20	GSE123496	Parietal cortex	2019	5	5	Illumina HiSeq 3000	Yes	Yes	Single-end
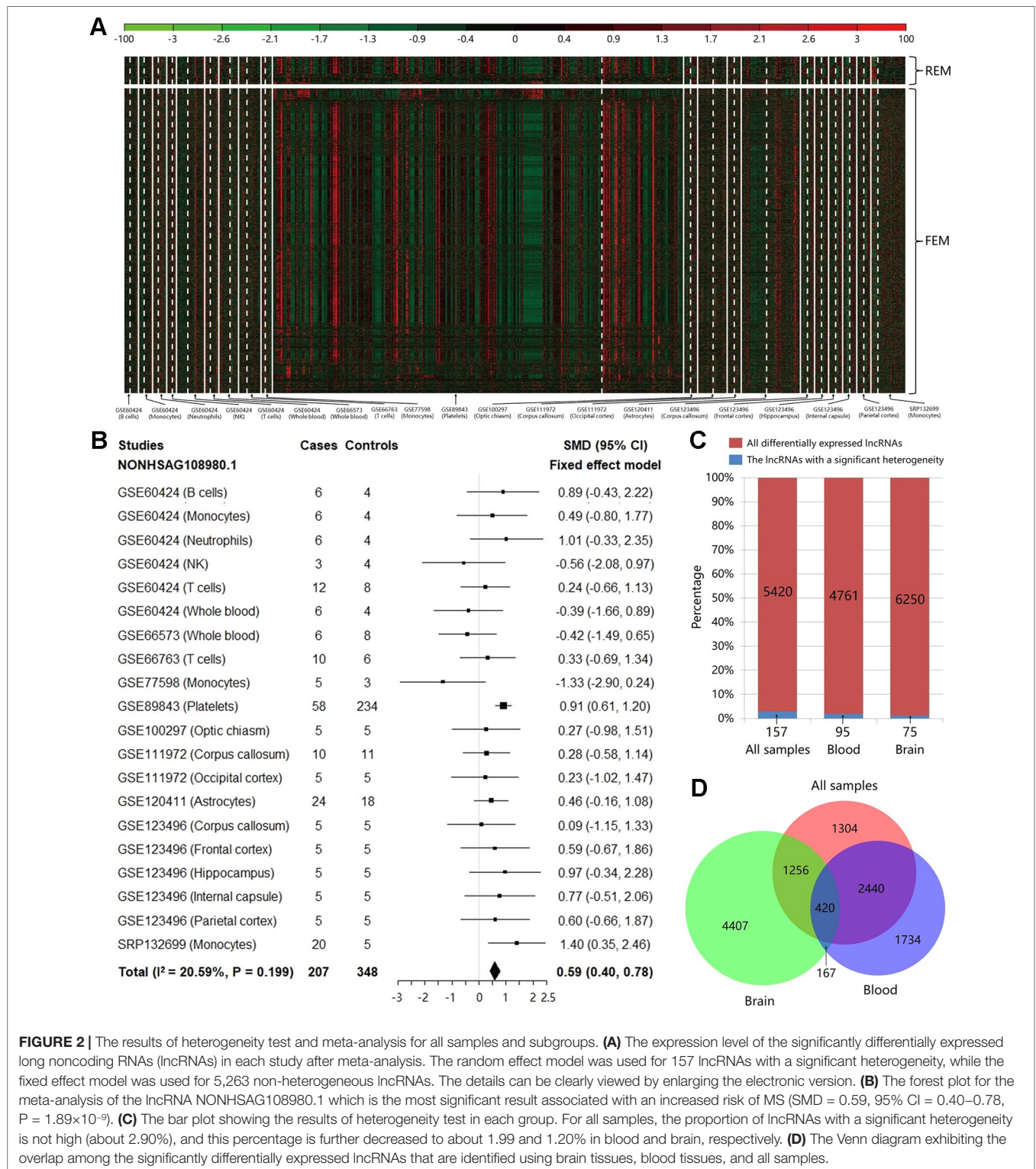
FIGURE 2 | The results of heterogeneity test and meta-analysis for all samples and subgroups. **(A)** The expression level of the significantly differentially expressed long noncoding RNAs (lncRNAs) in each study after meta-analysis. The random effect model was used for 157 lncRNAs with a significant heterogeneity, while the fixed effect model was used for 5,263 non-heterogeneous lncRNAs. The details can be clearly viewed by enlarging the electronic version. **(B)** The forest plot for the meta-analysis of the lncRNA NONHSAG108980.1 which is the most significant result associated with an increased risk of MS (SMD = 0.59, 95% CI = 0.40–0.78, $P = 1.89 \times 10^{-9}$). **(C)** The bar plot showing the results of heterogeneity test in each group. For all samples, the proportion of lncRNAs with a significant heterogeneity is not high (about 2.90%), and this percentage is further decreased to about 1.99 and 1.20% in blood and brain, respectively. **(D)** The Venn diagram exhibiting the overlap among the significantly differentially expressed lncRNAs that are identified using brain tissues, blood tissues, and all samples.

most significant association with an increased risk of MS (SMD = 0.59, 95% CI = 0.40–0.78, $P = 1.89 \times 10^{-9}$). Then, to investigate the heterogeneity of the lncRNA expression profile in various tissues, we split the samples into brain and blood tissue, and performed the heterogeneity test and meta-analysis for subgroups. We

found that not only the proportion of lncRNAs with a significant heterogeneity was not high for the whole samples, but also this percentage is further reduced to about 1.99 and 1.20% in blood and brain, respectively (**Figure 2C**). Finally, we explored the difference of the differentially expressed lncRNAs identified in

various tissues. We found that there was the higher specificity for these lncRNAs identified in brain compared with them identified in blood. Particularly, about 60.06% of the 5,420 differentially expressed lncRNAs can also be identified in the blood, while percentage is only 26.82% in brain (**Figure 2D**). Moreover, the total number of upregulated lncRNAs is far more than that of the downregulated ones in the blood (**Supplementary Table S2**) and the brain (**Supplementary Table S3**), which indicated that MS risk was related to lncRNA overexpression.

In addition, previous studies found that lncRNAs were modestly evolutionarily conserved in sequence (Guttman et al., 2009; Iyer et al., 2015). Therefore, we explored the conservation in sequence of these differentially expressed lncRNAs using conservation constrain search in NONCODE which contains the conservation information of lncRNAs in 13 common model organisms (i.e., human, chimp, gorilla, orangutan, rhesus, mouse, rat, cow, pig, opossum, platypus, chicken, and zebrafish). The results showed that only 0.11% of the differential lncRNAs were conserved in sequence among all these 13 organisms, while this percentage is increased to 28.5% in primates (human, chimp, gorilla, orangutan, and rhesus).

Tissue Specificity Analysis of the Multiple Sclerosis-Related Long Noncoding Ribonucleic Acids

Using expression data of NONCODE database, we performed the JS divergence metric and the cluster analysis to explore the tissue specificity of MS-related lncRNAs. The results of JS divergence metric showed that the MS-related lncRNA had high tissue specificity when used the brain, blood and whole samples (**Figure 3A**). For cluster analysis, relied on the same data, we further compared the expression patterns of these differentially expressed lncRNAs in various human tissues and cell lines. We found that the differentially expressed lncRNAs identified based on whole sample were highly specifically expressed in brain tissue (**Figure 3B**). Similarly, we observed a significant brain-specific expression for the differentially expressed lncRNAs identified based on brain sample (**Figure 3C**). Interestingly, although the differentially expressed lncRNAs were identified from blood sample, their expressions were still highly specific in brain tissue (**Figure 3D**). These results are consistent with the findings of the previous step and our recently published study (Han et al., 2019), which suggest that MS possesses the characteristics of the CNS disorder in lncRNA dysregulation.

Inferring the Functions of Multiple Sclerosis-Related Long Noncoding Ribonucleic Acids by Weighted Gene Co-Expression Network Analysis

After abundance quantification together with meta-analysis, we identified 2,051 protein-coding genes significantly differentially expressed between MS patients and controls (**Supplementary Table S4**). Then, we combined the count values of 2,051 differentially expressed protein-coding genes and 5,420 MS-related lncRNAs to perform the WGCNA. By quality control,

we removed three outlier samples whose minimum cluster size less than 5 and cutting height less than 4.0×10^6 (**Supplementary Figure S1**). The satisfactory soft threshold power β was set as 9 when the model fitting index R^2 equals 0.8 and the mean connectivity is close to 0 simultaneously (**Supplementary Figure S2**). Finally, we constructed a co-expression network which includes 1,938 protein-coding genes and 5,022 lncRNAs, and according to the interconnectedness of gene pairs, they were clustered into 15 modules in network (MEyellow, MEturquoise, MEblue, MEsalmon, MEred, MEpurple, MEpink, MEmagenta, MEgreen, MEmidnightblue, MEcyan, MEtan, MEgreenyellow, MEbrown, and MEblack) (**Figure 4A**). Moreover, to assess the significance of these modules for MS, we calculated two types of correlations as the index. The results of the average correlation of the genes in each module with the disease states showed that MEyellow is the most associated module with MS (r = 0.33, P = 5×10^{-15}), and the following three are MEred (r = 0.32, P = 2×10^{-14}), MEpink (r = −0.28, P = 2×10^{-11}), and MEbrown (r = 0.24, P = 9×10^{-9}). This was also applied to assess the association of each module with the platforms and the tissue types, respectively. Consistently, we found that the MEred (r = 0.71, P = 2×10^{-85}), MEbrown (r = 0.52, P = 1×10^{-39}), and MEyellow (r = 0.38, P = 2×10^{-20}) were most significantly associated with the tissue types. While there is no module strongly associated with platforms (**Figure 4B**). These findings are generally consistent with the result of the correlation between the module membership and the gene significance for MS. For example, MEyellow and MEred are the top two module with the high average correlation of genes with disease states, and they also show a high correlation between module membership and gene significance (cor = 0.43, P = 4.6×10^{-15} and cor = 0.50, P = 2.6×10^{-19}, respectively) (**Figures 4C, D**). On the contrary, MEcyan shows a very low level both for the two types of correlations (r = −0.058, P = 0.2 and cor = 0.038, P = 0.8) (**Figure 4E**).

In addition, we also perform a WGCNA with the satisfactory soft threshold power β = 9 using all the quantified genes. We found that these genes are clustered into 119 modules in the network, and about 82.2% differentially expressed genes are clustered into 16 of the 119 modules (including a gray one). We also found that these modules show low/modest association with MS (the correlation coefficients are < 0.19). These results reflect the similar distribution of the differentially expressed genes between using all and filtering genes in this WGCNA, and imply that the extra genes may mask the association of the differentially expressed genes with MS.

Pathway Analysis of the Multiple Sclerosis-Related Long Noncoding Ribonucleic Acids by Gene Set Enrichment Analysis

To explore the function of lncRNAs in MS, we performed GSEA in the four most significant modules for MS based on the two types of correlations, i.e., MEyellow (r = 0.33, P = 5×10^{-15} and cor = 0.43, P = 4.6×10^{-15}), MEred (r = 0.32, P = 2×10^{-14} and cor = 0.50, P = 2.6×10^{-19}), MEpink (r = −0.28, P = 2×10^{-11} and cor = 0.63, P = 3.5×10^{-14}), and MEbrown (r

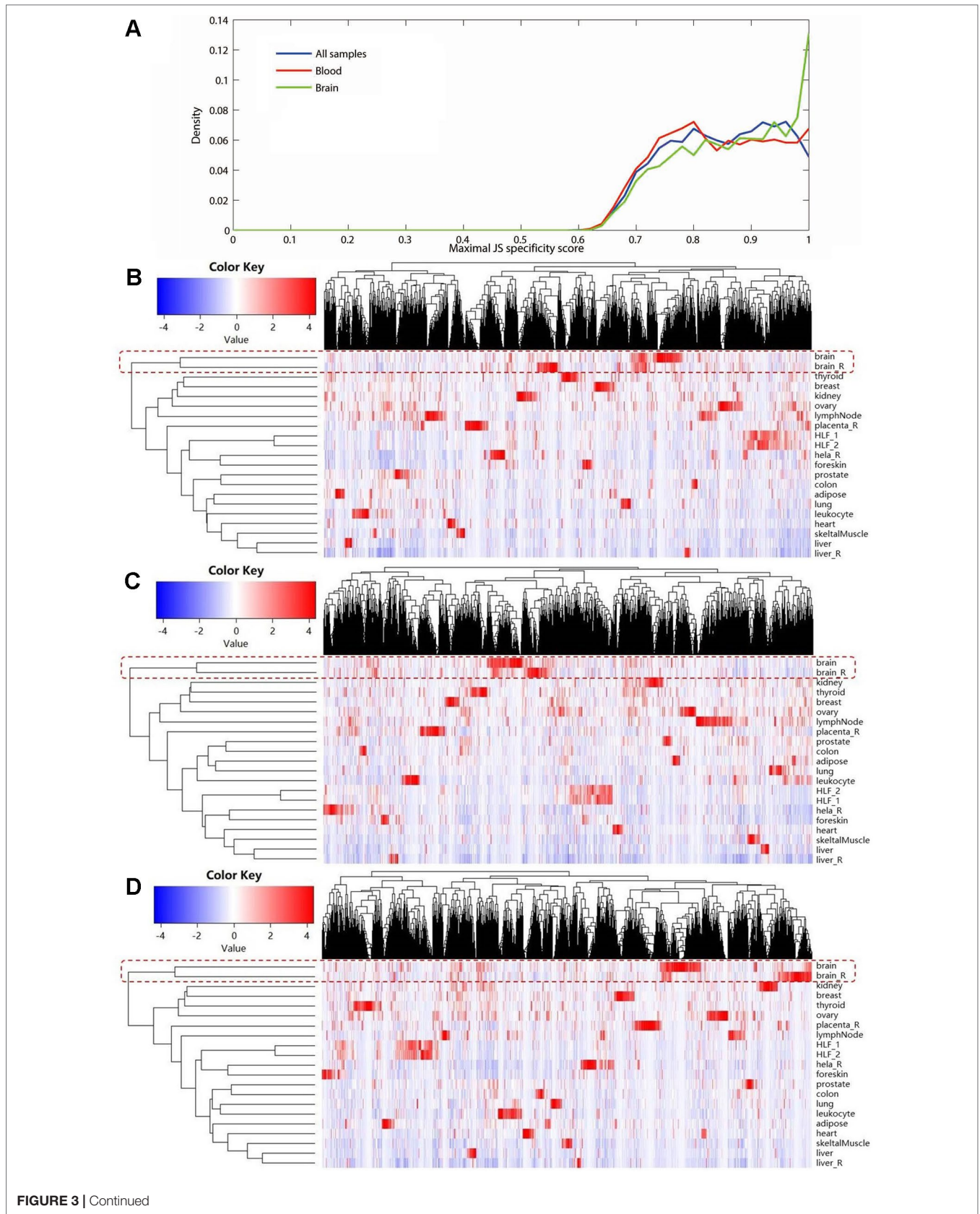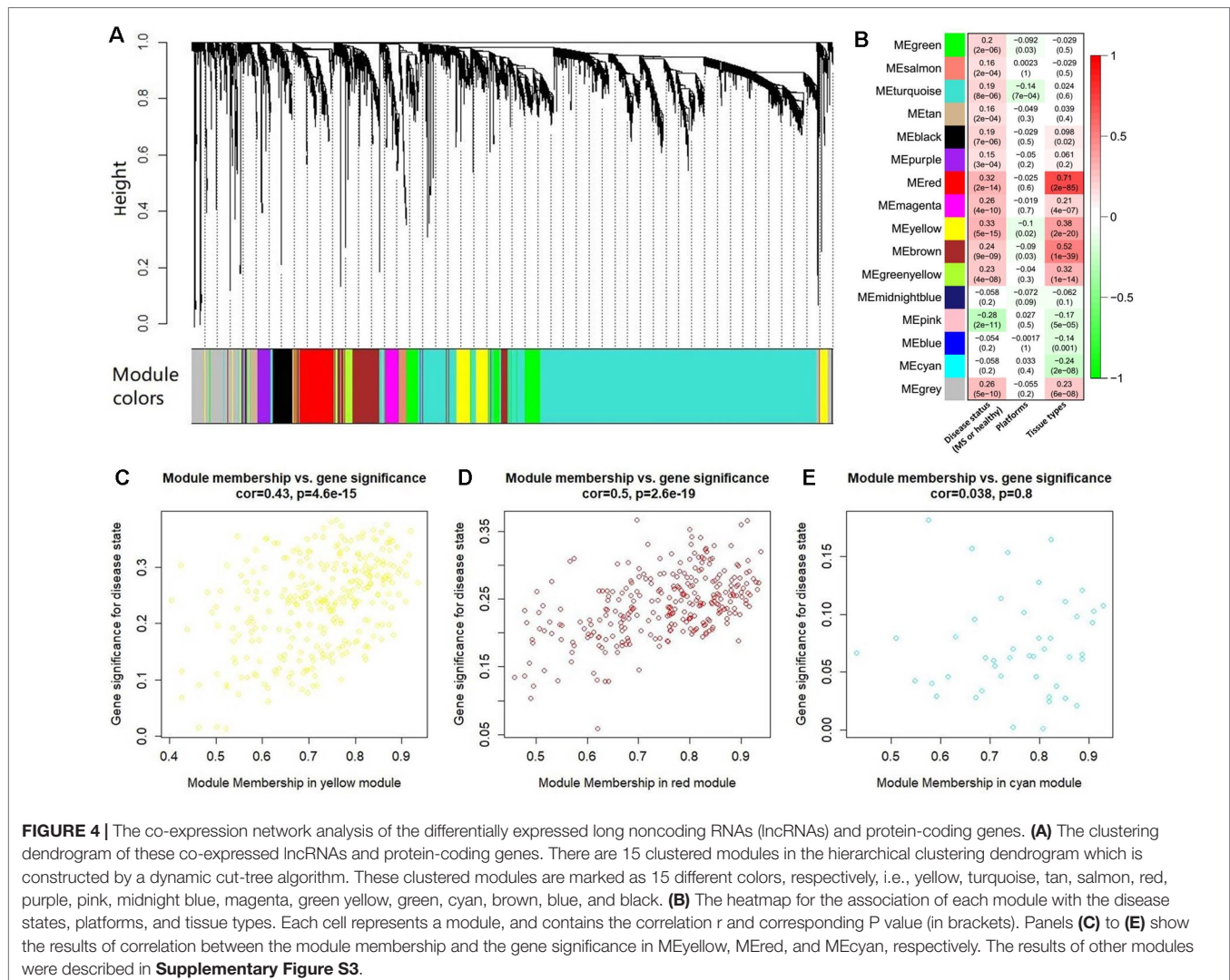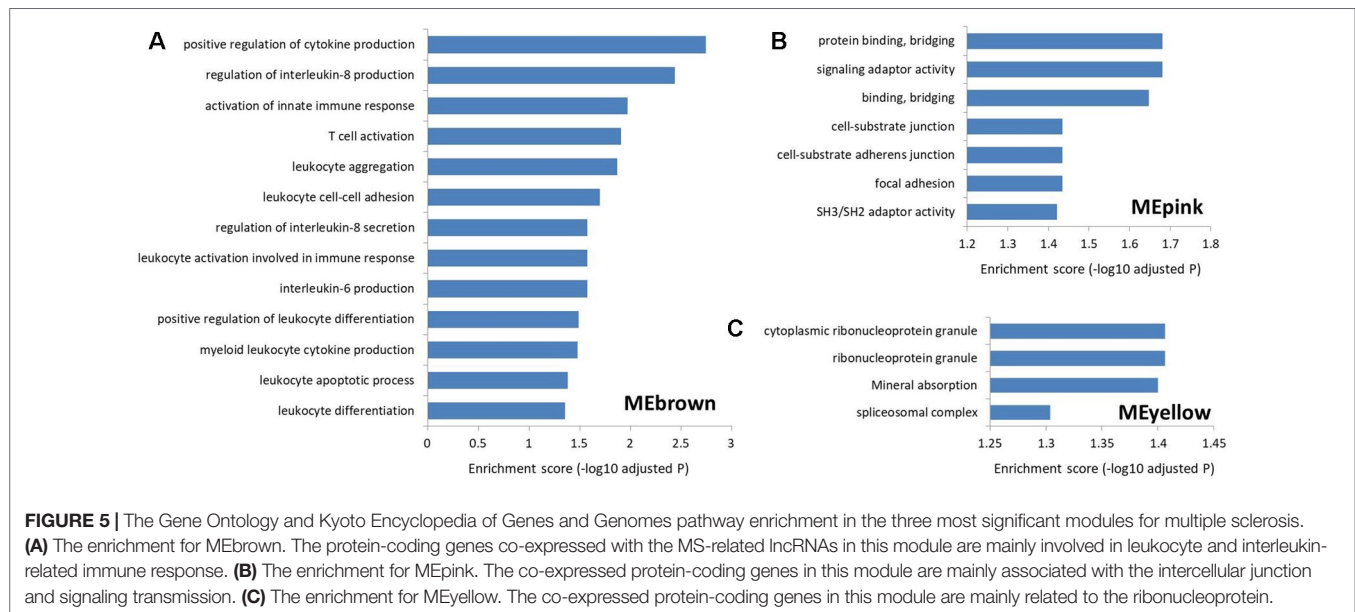**FIGURE 3 |** Continued

= 0.24, P = 9×10⁻⁹ and cor = 0.32, P = 4.7×10⁻⁹). We found no significantly enriched pathway related to the MEred. Based on the result of LncRNA2Target, we identified that two differentially expressed lncRNAs in MEred could target the MS-related genes. Particularly, two target genes (CDH1 and CDH2) of the lncRNA NONHSAG081583.2 encoded cadherin protein which is the most abundant adhesion molecules participating in nerve conduction in synaptic junctions and the proinflammatory cytokines in MS can downregulate its expression (Minagar et al., 2003; Tian et al., 2009). The lncRNA NONHSAG000840.2 targets a MS-related gene NOTCH2, and reducing NOTCH2 in the proinflammatory monocytes can increase the frequency

of the nonclassical monocytes and neutralizing antidrug antibody induction in IFN-β treated MS patients (Adriani et al., 2018). For MEbrown, the co-expressed protein-coding genes were mainly involved in leukocytes and interleukin-related immune response (**Figure 5A** and **Supplementary Table S5**), which was similar to the finding of our recent study (Han et al., 2019). Many genomic variants in the human leukocyte antigen complexes and interleukin receptor were identified significantly associated with susceptibility of MS (Rubio et al., 2002; Teutsch et al., 2003; Lundmark et al., 2007; Hollenbach and Oksenberg, 2015; Tang et al., 2019a). The protein-coding genes in MEpink are mainly associated with intercellular junction

FIGURE 4 | The co-expression network analysis of the differentially expressed long noncoding RNAs (lncRNAs) and protein-coding genes. (A) The clustering dendrogram of these co-expressed lncRNAs and protein-coding genes. There are 15 clustered modules in the hierarchical clustering dendrogram which is constructed by a dynamic cut-tree algorithm. These clustered modules are marked as 15 different colors, respectively, i.e., yellow, turquoise, tan, salmon, red, purple, pink, midnight blue, magenta, green yellow, green, cyan, brown, blue, and black. (B) The heatmap for the association of each module with the disease states, platforms, and tissue types. Each cell represents a module, and contains the correlation r and corresponding P value (in brackets). Panels (C) to (E) show the results of correlation between the module membership and the gene significance in MEyellow, MEred, and MEcyan, respectively. The results of other modules were described in **Supplementary Figure S3**.

FIGURE 5 | The Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway enrichment in the three most significant modules for multiple sclerosis. **(A)** The enrichment for MEbrown. The protein-coding genes co-expressed with the MS-related lncRNAs in this module are mainly involved in leukocyte and interleukin-related immune response. **(B)** The enrichment for MEpink. The co-expressed protein-coding genes in this module are mainly associated with the intercellular junction and signaling transmission. **(C)** The enrichment for MEyellow. The co-expressed protein-coding genes in this module are mainly related to the ribonucleoprotein.

and signaling transmission (**Figure 5B**). Previous studies found that the defect of axon-glial signaling transmission caused by the oligodendrocyte gap junction loss and disconnection contributes to MS pathogenesis (Brand-Schieber et al., 2005; Markoullis et al., 2012; Markoullis et al., 2014). The results of LncRNA2Target showed that lncRNA NONHSAG049754.2 in MEyellow targets the MS-related gene TNFRSF10A. This gene encodes the receptor of tumor necrosis factor (TNF) cytokines which plays a important role in inflammation regulations and is related to susceptibility of developing MS (De-la-Torre et al., 2019). The protein-coding genes in the MEyellow are related to ribonucleoprotein (**Figure 5C**). Ribonucleoprotein is a kind of ribonucleic acid-binding protein which participates in the mRNA splicing (Guthrie, 1991). Previous study showed that as an important autoantigen in the neuroimmune disease, the ribonucleoprotein significantly more often interact with the autoantibodies in MS cerebrospinal fluids compared with controls (Sueoka et al., 2004; Yukitake et al., 2008). The following studies further identified a ribonucleoprotein-related lncRNA, TNF-α, and heterogeneous nuclear ribonucleoprotein L, which was significantly upregulated and produced transcriptional activating complexes to promote TNF-α expression by cooperating with ribonucleoprotein in the circulating blood cells of MS (Li et al., 2014; Eftekharian et al., 2017). Given that MEyellow is the most significant module for MS, we inferred that one of the key mechanisms of lncRNAs in MS is associated with the regulation of ribonucleoprotein and TNF cytokines receptor.

CONCLUSIONS

In this study, we comprehensively collected MS-related RNA-seq data from a variety of studies, and integrated these data by an expression-based meta-analysis to assess the affection of lncRNAs on the MS pathogenesis on genome scale. We identified a total of 5,420 lncRNAs significantly differentially expressed between

MS patients and controls. Then, the subgroup analysis found a small heterogeneity of the lncRNA expression profile between the brain and blood tissues. Further, the specificity analysis of multiple tissues showed that the differentially expressed lncRNAs (including identified using brain, blood, and whole sample) are highly specifically expressed in brain tissue. Finally, the result of GSEA and WGCNA demonstrated that the potential important function of lncRNAs in MS may be involved in the regulation of ribonucleoprotein and TNF cytokines receptor. All in all, we performed a strategy to resolve the inconsistent MS-related lncRNA findings in previous studies, and explore the functions of these lncRNAs in MS. The findings of this study will be benefit to improve the understanding of the pathogenesis of MS.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/gds, https://www.ebi.ac.uk/arrayexpress, https://ddbj.nig.ac.jp/DRASearch.

AUTHOR CONTRIBUTIONS

ZH and FZ designed the research. ZH, FZ, JH, and WX collected the data. ZH performed the research, analyzed data, and wrote the paper. FZ reviewed and modified the manuscript. All authors discussed the results, and contributed to the final manuscript. All authors read and approved the final manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2019.01136/full#supplementary-material

REFERENCES

Adriani, M., Nytrova, P., Mbogning, C., Hassler, S., Medek, K., Jensen, P. E. H., et al. (2018). Monocyte NOTCH2 expression predicts IFN-beta immunogenicity in multiple sclerosis patients. *JCI Insight* 3. doi: 10.1172/jci.insight.99274

Aken, B. L., Achuthan, P., Akanni, W., Amode, M. R., Bernsdorff, F., Bhai, J., et al. (2017). Ensembl 2017. *Nucleic Acids Res.* 45, D635–D642. doi: 10.1093/nar/gkw1104

Angerer, I. C., Hecker, M., Koczan, D., Roch, L., Friess, J., Ruge, A., et al. (2018). Transcriptome profiling of peripheral blood immune cell populations in multiple sclerosis patients before and during treatment with a sphingosine-1-phosphate receptor modulator. *CNS Neurosci. Ther.* 24, 193–201. doi: 10.1111/cns.12793

Athar, A., Fullgrabe, A., George, N., Iqbal, H., Huerta, L., Ali, A., et al. (2019). ArrayExpress update - from bulk to single-cell expression data. *Nucleic Acids Res.* 47, D711–D715. doi: 10.1093/nar/gky964

Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.* 41, D991–D995. doi: 10.1093/nar/gks1193

Brand-Schieber, E., Werner, P., Iacobas, D. A., Iacobas, S., Beelitz, M., Lowery, S. L., et al. (2005). Connexin43, the major gap junction protein of astrocytes, is down-regulated in inflamed white matter in an animal model of multiple sclerosis. *J. Neurosci. Res.* 80, 798–808. doi: 10.1002/jnr.20474

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527. doi: 10.1038/nbt.3519

Browne, P., Chandraratna, D., Angood, C., Tremlett, H., Baker, C., Taylor, B. V., et al. (2014). Atlas of multiple sclerosis 2013: a growing global problem with widespread inequity. *Neurology* 83, 1022–1024. doi: 10.1212/WNL.0000000000000768

Bullard, J. H., Purdom, E., Hansen, K. D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinf.* 11, 94. doi: 10.1186/1471-2105-11-94

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., et al. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–1927. doi: 10.1101/gad.17446611

Cardamone, G., Paraboschi, E. M., Solda, G., Cantoni, C., Supino, D., Piccio, L., et al. (2019). Not only cancer: the long non-coding RNA MALAT1 affects the repertoire of alternatively spliced transcripts and circular RNAs in multiple sclerosis. *Hum. Mol. Genet.* 28, 1414–1428. doi: 10.1093/hmg/ddy438

Chen, Y. G., Satpathy, A. T., and Chang, H. Y. (2017). Gene regulation in the immune system by long noncoding RNAs. *Nat. Immunol.* 18, 962–972. doi: 10.1038/ni.3771

Cheng, L., and Hu, Y. (2018). Human disease system biology. *Curr. Gene Ther.* doi: 10.2174/1566523218666181101143116

Cheng, L., Hu, Y., Sun, J., Zhou, M., and Jiang, Q. (2018). DincRNA: a comprehensive web-based bioinformatics toolkit for exploring disease associations and ncRNA function. *Bioinformatics* 34, 1953–1956. doi: 10.1093/bioinformatics/bty002

Cheng, L., Sun, J., Xu, W., Dong, L., Hu, Y., and Zhou, M. (2016). OAHG: an integrated resource for annotating human genes with multi-level ontologies. *Sci. Rep.* 6, 34820. doi: 10.1038/srep34820

Cheng, L., Wang, P., Tian, R., Wang, S., Guo, Q., Luo, M., et al. (2019). LncRNA2Target v2.0: a comprehensive database for target genes of lncRNAs in human and mouse. *Nucleic Acids Res.* 47, D140–D144. doi: 10.1093/nar/gky1051

Church, D. M., Schneider, V. A., Graves, T., Auger, K., Cunningham, F., Bouk, N., et al. (2011). Modernizing reference genome assemblies. *PloS Biol.* 9, e1001091. doi: 10.1371/journal.pbio.1001091

De-la-Torre, A., Silva-Aldana, C. T., Munoz-Ortiz, J., Pineros-Hernandez, L. B., Otero, O., Dominguez, A., et al. (2019). Uveitis and multiple sclerosis: potential common causal mutations. *Mol. Neurobiol.* doi: 10.1007/s12035-019-1630-2

Dong, X., Chen, K., Cuevas-Diaz Duran, R., You, Y., Sloan, S. A., Zhang, Y., et al. (2015). Comprehensive identification of long non-coding RNAs in purified cell types from the brain reveals functional LncRNA in OPC fate determination. *PloS Genet.* 11, e1005669. doi: 10.1371/journal.pgen.1005669

Du, Z., Fei, T., Verhaak, R. G., Su, Z., Zhang, Y., Brown, M., et al. (2013). Integrative genomic analyses reveal clinically relevant long noncoding RNAs in human cancer. *Nat. Struct. Mol. Biol.* 20, 908–913. doi: 10.1038/nsmb.2591

Eftekharian, M. M., Ghafouri-Fard, S., Soudyab, M., Omrani, M. D., Rahimi, M., Sayad, A., et al. (2017). Expression analysis of long non-coding RNAs in the blood of multiple sclerosis patients. *J. Mol. Neurosci.* 63, 333–341. doi: 10.1007/s12031-017-0982-1

Fang, S., Zhang, L., Guo, J., Niu, Y., Wu, Y., Li, H., et al. (2018). NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res.* 46, D308–D314. doi: 10.1093/nar/gkx1107

Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev. Genet.* 15, 7–21. doi: 10.1038/nrg3606

Frohman, E. M., Racke, M. K., and Raine, C. S. (2006). Multiple sclerosis–the plaque and its pathogenesis. *N Engl. J. Med.* 354, 942–955. doi: 10.1056/NEJMra052130

Gellert, P., Ponomareva, Y., Braun, T., and Uchida, S. (2013). Noncoder: a web interface for exon array-based detection of long non-coding RNAs. *Nucleic Acids Res.* 41, e20. doi: 10.1093/nar/gks877

Gharesouran, J., Taheri, M., Sayad, A., Ghafouri-Fard, S., Mazdeh, M., and Omrani, M. D. (2019). A novel regulatory function of long non-coding RNAs at different levels of gene expression in multiple sclerosis. *J. Mol. Neurosci.* 67, 434–440. doi: 10.1007/s12031-018-1248-2

Gomez, J. A., Wapinski, O. L., Yang, Y. W., Bureau, J. F., Gopinath, S., Monack, D. M., et al. (2013). The NeST long ncRNA controls microbial susceptibility and epigenetic activation of the interferon-gamma locus. *Cell* 152, 743–754. doi: 10.1016/j.cell.2013.01.015

Guthrie, C. (1991). Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. *Science* 253, 157–163. doi: 10.1126/science.1853200

Guttman, M., Amit, I., Garber, M., French, C., Lin, M. F., Feldser, D., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458, 223–227. doi: 10.1038/nature07672

Han, Z., Jiang, Q., Zhang, T., Wu, X., Ma, R., Wang, J., et al. (2015). Analyzing large-scale samples confirms the association between the rs1051730 polymorphism and lung cancer susceptibility. *Sci. Rep.* 5, 15642. doi: 10.1038/srep15642

Han, Z., Qu, J., Zhao, J., and Zou, X. (2018a). Analyzing 74,248 samples confirms the association between CLU rs11136000 polymorphism and Alzheimer's disease in caucasian but not Chinese population. *Sci. Rep.* 8, 11062. doi: 10.1038/s41598-018-29450-2

Han, Z., Qu, J., Zhao, J., and Zou, X. (2018b). Genetic variant rs755622 regulates expression of the multiple sclerosis severity modifier D-dopachrome tautomerase in a sex-specific Way. *BioMed. Res. Int.* 2018, 8285653. doi: 10.1155/2018/8285653

Han, Z., Xue, W., Tao, L., Lou, Y., Qiu, Y., and Zhu, F. (2019). Genome-wide identification and analysis of the eQTL lncRNAs in multiple sclerosis based on RNA-seq data. *Brief Bioinform.* doi: 10.1093/bib/bbz036

He, D., Wang, J., Lu, Y., Deng, Y., Zhao, C., Xu, L., et al. (2017). lncRNA functional networks in oligodendrocytes reveal stage-specific myelination control by an lncOL1/Suz12 complex in the CNS. *Neuron* 93, 362–378. doi: 10.1016/j.neuron.2016.11.044

Hollenbach, J. A., and Oksenberg, J. R. (2015). The immunogenetics of multiple sclerosis: a comprehensive review. *J. Autoimmun.* 64, 13–25. doi: 10.1016/j.jaut.2015.06.010

Ilott, N. E., and Ponting, C. P. (2013). Predicting long non-coding RNAs using RNA sequencing. *Methods* 63, 50–59. doi: 10.1016/j.ymeth.2013.03.019

Iyer, M. K., Niknafs, Y. S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* 47, 199–208. doi: 10.1038/ng.3192

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. doi: 10.1093/nar/gkw1092

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., et al. (2002). The human genome browser at UCSC. *Genome Res.* 12, 996–1006. doi: 10.1101/gr.229102

Kim, H., Kim, J. H., Kim, S. Y., Jo, D., Park, H. J., Kim, J., et al. (2015). Meta-analysis of large-scale toxicogenomic data finds neuronal regeneration related protein and cathepsin D to be novel biomarkers of drug-induced toxicity. *PloS One* 10, e0136698. doi: 10.1371/journal.pone.0136698

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9, 559. doi: 10.1186/1471-2105-9-559

Leinonen, R., Sugawara, H., and Shumway, M.International Nucleotide Sequence Database, C. (2011). The sequence read archive. *Nucleic Acids Res.* 39, D19–D21. doi: 10.1093/nar/gkq1019

Li, B., Tang, J., Yang, Q., Li, S., Cui, X., Li, Y., et al. (2017). NOREVA: normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Res.* 45, W162–W170. doi: 10.1093/nar/gkx449

Li, Y., Song, D., Jiang, Y., Wang, J., Feng, R., Zhang, L., et al. (2016). CR1 rs3818361 Polymorphism contributes to Alzheimer's disease susceptibility in Chinese population. *Mol. Neurobiol.* 53, 4054–4059. doi: 10.1007/s12035-015-9343-7

Li, Y. H., Li, X. X., Hong, J. J., Wang, Y. X., Fu, J. B., Yang, H., et al. (2019). Clinical trials, progression-speed differentiating features and swiftness rule of the innovative targets of first-in-class drugs. *Brief Bioinform.* doi: 10.1093/bib/bby130

Li, Y. H., Yu, C. Y., Li, X. X., Zhang, P., Tang, J., Yang, Q., et al. (2018). Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res.* 46, D1121–D1127. doi: 10.1093/nar/gkx1076

Li, Z., Chao, T. C., Chang, K. Y., Lin, N., Patil, V. S., Shimizu, C., et al. (2014). The long noncoding RNA THRIL regulates TNFalpha expression through its interaction with hnRNPL. *Proc. Natl. Acad. Sci. U.S.A.* 111, 1002–1007. doi: 10.1073/pnas.1313768111

Liu, G., Xu, Y., Jiang, Y., Zhang, L., Feng, R., and Jiang, Q. (2017). PICALM rs3851179 variant confers susceptibility to Alzheimer's disease in Chinese population. *Mol. Neurobiol.* 54, 3131–3136. doi: 10.1007/s12035-016-9886-2

Lundmark, F., Duvefelt, K., Iacobaeus, E., Kockum, I., Wallstrom, E., Khademi, M., et al. (2007). Variation in interleukin 7 receptor alpha chain (IL7R) influences risk of multiple sclerosis. *Nat. Genet.* 39, 1108–1113. doi: 10.1038/ng2106

Ma, T., Huo, Z., Kuo, A., Zhu, L., Fang, Z., Zeng, X., et al. (2019). MetaOmics: analysis pipeline and browser-based software suite for transcriptomic meta-analysis. *Bioinformatics* 35, 1597–1599. doi: 10.1093/bioinformatics/bty825

Markoullis, K., Sargiannidou, I., Schiza, N., Hadjisavvas, A., Roncaroli, F., Reynolds, R., et al. (2012). Gap junction pathology in multiple sclerosis lesions and normal-appearing white matter. *Acta Neuropathol.* 123, 873–886. doi: 10.1007/s00401-012-0978-4

Markoullis, K., Sargiannidou, I., Schiza, N., Roncaroli, F., Reynolds, R., and Kleopa, K. A. (2014). Oligodendrocyte gap junction loss and disconnection from reactive astrocytes in multiple sclerosis gray matter. *J. Neuropathol. Exp. Neurol.* 73, 865–879. doi: 10.1097/NEN.0000000000000106

Masoumi, F., Ghorbani, S., Talebi, F., Branton, W. G., Rajaei, S., Power, C., et al. (2019). Malat1 long noncoding RNA regulates inflammation and leukocyte differentiation in experimental autoimmune encephalomyelitis. *J. Neuroimmunol.* 328, 50–59. doi: 10.1016/j.jneuroim.2018.11.013

Minagar, A., Ostanin, D., Long, A. C., Jennings, M., Kelley, R. E., Sasaki, M., et al. (2003). Serum from patients with multiple sclerosis downregulates occludin and VE-cadherin expression in cultured endothelial cells. *Mult. Scler* 9, 235–238. doi: 10.1191/1352458503ms916oa

Ng, S. Y., Lin, L., Soh, B. S., and Stanton, L. W. (2013). Long noncoding RNAs in development and disease of the central nervous system. *Trends Genet.* 29, 461–468. doi: 10.1016/j.tig.2013.03.002

Ogasawara, O., Mashima, J., Kodama, Y., Kaminuma, E., Nakamura, Y., Okubo, K., et al. (2013). DDBJ new system and service refactoring. *Nucleic Acids Res.* 41, D25–D29. doi: 10.1093/nar/gks1152

Pahlevan Kakhki, M., Nikravesh, A., Shirvani Farsani, Z., Sahraian, M. A., and Behmanesh, M. (2018). HOTAIR but not ANRIL long non-coding RNA contributes to the pathogenesis of multiple sclerosis. *Immunology* 153, 479–487. doi: 10.1111/imm.12850

Rezazadeh, M., Gharesouran, J., Moradi, M., Noroozi, R., Omrani, M. D., Taheri, M., et al. (2018). Association study of ANRIL genetic variants and multiple sclerosis. *J. Mol. Neurosci.* 65, 54–59. doi: 10.1007/s12031-018-1069-3

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Rubio, J. P., Bahlo, M., Butzkueven, H., van Der Mei, I. A., Sale, M. M., Dickinson, J. L., et al. (2002). Genetic dissection of the human leukocyte antigen region by use of haplotypes of Tasmanians with multiple sclerosis. *Am. J. Hum. Genet.* 70, 1125–1137. doi: 10.1086/339932

Santoro, M., Nociti, V., Lucchini, M., De Fino, C., Losavio, F. A., and Mirabella, M. (2016). Expression profile of long non-coding RNAs in serum of patients with multiple sclerosis. *J. Mol. Neurosci.* 59, 18–23. doi: 10.1007/s12031-016-0741-8

Sayers, E. W., Agarwala, R., Bolton, E. E., Brister, J. R., Canese, K., Clark, K., et al. (2019). Database resources of the national center for biotechnology Information. *Nucleic Acids Res.* 47, D23–D28. doi: 10.1093/nar/gky1069

Selmaj, I., Cichalewska, M., Namiecinska, M., Galazka, G., Horzelski, W., Selmaj, K. W., et al. (2017). Global exosome transcriptome profiling reveals biomarkers for multiple sclerosis. *Ann. Neurol.* 81, 703–717. doi: 10.1002/ana.24931

Soneson, C., Love, M. I., and Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* 4, 1521. doi: 10.12688/f1000research.7563.2

Sospedra, M., and Martin, R. (2005). Immunology of multiple sclerosis. *Annu. Rev. Immunol.* 23, 683–747. doi: 10.1146/annurev.immunol.23.021704.115707

Sueoka, E., Yukitake, M., Iwanaga, K., Sueoka, N., Aihara, T., and Kuroda, Y. (2004). Autoantibodies against heterogeneous nuclear ribonucleoprotein B1 in CSF of MS patients. *Ann. Neurol.* 56, 778–786. doi: 10.1002/ana.20276

Szajewska, H., and Kolodziej, M. (2015). Systematic review with meta-analysis: *Saccharomyces boulardii* in the prevention of antibiotic-associated diarrhoea. *Aliment Pharmacol. Ther.* 42, 793–801. doi: 10.1111/apt.13344

Tang, J., Fu, J., Wang, Y., Li, B., Li, Y., Yang, Q., et al. (2019a). ANPELA: analysis and performance assessment of the label-free quantification workflow for metaproteomic studies. *Brief Bioinform.* doi: 10.1093/bib/bby127

Tang, J., Fu, J., Wang, Y., Luo, Y., Yang, Q., Li, B., et al. (2019b). Simultaneous improvement in the precision, accuracy, and robustness of label-free proteome quantification by optimizing data manipulation chains. *Mol. Cell Proteomics* 18, 1683–1699. doi: 10.1074/mcp.RA118.001169

Teutsch, S. M., Booth, D. R., Bennetts, B. H., Heard, R. N., and Stewart, G. J. (2003). Identification of 11 novel and common single nucleotide polymorphisms in the interleukin-7 receptor-alpha gene and their associations with multiple sclerosis. *Eur. J. Hum. Genet.* 11, 509–515. doi: 10.1038/sj.ejhg.5200994

The Gene Ontology, C. (2017). Expansion of the gene ontology knowledgebase and resources. *Nucleic Acids Res.* 45, D331–D338. doi: 10.1093/nar/gkw1108

Tian, L., Rauvala, H., and Gahmberg, C. G. (2009). Neuronal regulation of immune responses in the central nervous system. *Trends Immunol.* 30, 91–99. doi: 10.1016/j.it.2008.11.002

UniProt, C. (2015). UniProt: a hub for protein information. *Nucleic Acids Res.* 43, D204–D212. doi: 10.1093/nar/gku989

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484

Xue, W., Yang, F., Wang, P., Zheng, G., Chen, Y., Yao, X., et al. (2018). What contributes to serotonin-norepinephrine reuptake iinhibitors' dual-targeting mechanism? The key role of transmembrane domain 6 in human serotonin and norepinephrine transporters revealed by molecular dynamics simulation. *ACS Chem. Neurosci.* 9, 1128–1140. doi: 10.1021/acschemneuro.7b00490

Yang, Q., Li, B., Tang, J., Cui, X., Wang, Y., Li, X., et al. (2019). Consistent gene signature of schizophrenia identified by a novel feature selection strategy from comprehensive sets of transcriptomic data. *Brief Bioinform.* doi: 10.1093/bib/bbz049

Yin, J., Sun, W., Li, F., Hong, J., Li, X., Zhou, Y., et al. (2019). VARIDT 1.0: variability of drug transporter database. *Nucleic Acids Res*. doi: 10.1093/nar/gkz779

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118

Yukitake, M., Sueoka, E., Sueoka-Aragane, N., Sato, A., Ohashi, H., Yakushiji, Y., et al. (2008). Significantly increased antibody response to heterogeneous nuclear ribonucleoproteins in cerebrospinal fluid of multiple sclerosis patients but not in patients with human T-lymphotropic virus type I-associated myelopathy/tropical spastic paraparesis. *J. Neurovirol*. 14, 130–135. doi: 10.1080/13550280701883840

Zhang, F., Gao, C., Ma, X. F., Peng, X. L., Zhang, R. X., Kong, D. X., et al. (2016). Expression profile of long noncoding RNAs in peripheral blood mononuclear cells from multiple sclerosis patients. *CNS Neurosci. Ther*. 22, 298–305. doi: 10.1111/cns.12498

Zhang, M., Chang, Y.-C., Shankara, S., Jacobs, A., Godin, J., Klinger, K., et al. (2019). Characterization of the Peripheral Blood Transcriptome in Alemtuzumab-Treated Relapsing-Remitting Multiple Sclerosis Patients From the CARE-MS I and II Studies (P4. 2-041). AAN Enterprises.