scientific reports

OPEN

Genetic risk score in multiple sclerosis is associated with unique gut microbiome

Noha S. Elsayed[1,2], Robert K. Valenzuela[1], Terrie Kitchner[1], Thao Le[3], John Mayer[4], Zheng-Zheng Tang[5], Vishnu R. Bayanagari[1,6], Qiongshi Lu[5], Paula Aston[7], Karthik Anantharaman[8] & Sanjay K. Shukla[1,9✉]

Multiple sclerosis (MS) is a complex autoimmune disease in which both the roles of genetic susceptibility and environmental/microbial factors have been investigated. More than 200 genetic susceptibility variants have been identified along with the dysbiosis of gut microbiota, both independently have been shown to be associated with MS. We hypothesize that MS patients harboring genetic susceptibility variants along with gut microbiome dysbiosis are at a greater risk of exhibiting the disease. We investigated the genetic risk score for MS in conjunction with gut microbiota in the same cohort of 117 relapsing remitting MS (RRMS) and 26 healthy controls. DNA samples were genotyped using Illumina's Infinium Immuno array-24 v2 chip followed by calculating genetic risk score and the microbiota was determined by sequencing the V4 hypervariable region of the 16S rRNA gene. We identified two clusters of MS patients, Cluster A and B, both having a higher genetic risk score than the control group. However, the MS cases in cluster B not only had a higher genetic risk score but also showed a distinct gut microbiome than that of cluster A. Interestingly, cluster A which included both healthy control and MS cases had similar gut microbiome composition. This could be due to (i) the non-active state of the disease in that group of MS patients at the time of fecal sample collection and/or (ii) the restoration of the gut microbiome post disease modifying therapy to treat the MS. Our study showed that there seems to be an association between genetic risk score and gut microbiome dysbiosis in triggering the disease in a small cohort of MS patients. The MS Cluster A who have a higher genetic risk score but microbiome profile similar to that of healthy controls could be due to the remitting phase of the disease or due to the effect of disease modifying therapies.

Multiple sclerosis (MS), a disease that affects nearly 2.8 million people worldwide[1], is a chronic, inflammatory, autoimmune disease of the central nervous system with a complex, multifactorial etiology[2]. The symptoms of MS range from fatigue, numbness, muscle spasms and weakness to various gastrointestinal and urinary malfunction symptoms[3]. Pathologically, the disease manifests with demyelination and degeneration of neurons, and presence of white matter lesions on the brain and the spinal cord[1,3]. What etiological factors drives the two phenotypes of MS : relapsing–remitting MS (RRMS) and primary progressive MS (PPMS)[3] is not fully understood. The most common phenotype is the RRMS where the patient alternates between active and non-active episodes of symptoms. The active episodes are marked with motor, sensory and cognitive symptoms in addition to brain lesions detected by magnetic resonance imaging[4]. The complex etiology of MS disease continues to being investigated through increasing understanding of genetic susceptibility and different triggering modalities arising from life-style and/or environment[5] such as smoking, low sun light exposure, high salt diet, viral infection(s), and microbe(s) or microbial metabolites emanating from gut microbiome dysbiosis[6–9]. The gut microbiome with

[1]Center for Precision Medicine Research, Marshfield Clinic Research Institute, 1000 N Oak Avenue # MLR, Marshfield, WI 54449, USA. [2]Present address: Department of Pediatrics, Albert Einstein Medical College, New York, United States. [3]Integrated Research Development Laboratory, Marshfield Clinic Research Institute, Marshfield Clinic Health System, Marshfield, WI 54449, USA. [4]Office of Research Computing and Analytics, Marshfield Clinic Research Institute, Marshfield Clinic Health System, Marshfield, WI 54449, USA. [5]Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI, USA. [6]Present address: Roger Williams Medical Center, Boston University School of Medicine, Providence, RI 02908, USA. [7]Department of Neurology, Marshfield Clinic Health System, Marshfield, WI 54449, USA. [8]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI 53706, USA. [9]Computational and Informatics in Biology and Medicine Program, University of Wisconsin-Madison, Madison, WI 53706, USA. ✉email: shukla.sanjay@mcrf.mfldclin.edu

its dynamic reservoir of trillions of microbes representing hundreds of species is of great interest to potentially link its role in genetically susceptible persons.

Indeed, genetic susceptibility to MS is complex and hundreds of genomic regions that are implicated are dispersed throughout the chromosomes[10]. The genetic susceptibility accounts for 30% of the MS cases[11]. Siblings of MS patients are seven times more susceptible for this disease than general population[12]. However, the major histocompatibility complex (MHC) haplotypes on chromosome 6 have shown as the highest reproducible associations with MS susceptibility. Mostly MHC class II alleles, such as DQA1*01:01-DRB1*15:01 and DQB1*03:01-DQB1*03:02 are pivotal[13]. The strongest risk allele is HLA-DRB1*15:01 with an odds ratio of 3.08[14]. Usually, in complex diseases like MS, the more risk alleles the subject carries, the higher the predisposition to the disease[15]. Thus, measuring the genetic risk score of the MS patients in comparison to the healthy controls can be used in revealing more precise genetic susceptibility to this complex disease.

While knowledge of genetic susceptibility to MS has enhanced our understanding of the disease, the precise source and role of the environmental factor(s) including microbial trigger(s) associated with MS is far from settled. The human gut microbiome with its rich source of microbial diversity, their antigens, and metabolites are being explored as a possible source of infectious triggers modulating the MS disease. Indeed, several recent studies have reported association of gut microbiome dysbiosis with the MS[2,6,16]. A convincing role for the gut microbiome in the MS disease was supported by observation from an experimental autoimmune encephalomyelitis (EAE) disease mouse model analogous to MS. In this model, SJL/J mice were protected from MS when grown in germ free conditions[17]. Furthermore, their susceptibility to EAE was restored by exposing these mice to the commensal bacteria from fecal material from specific pathogen-free mice[17]. Additionally, the MS disease development was reproducible in a EAE mice model by transferring MS patient's fecal material to mice[5]. Since then, several studies have reported an association of MS with the gut microbiome dysbiosis involving different taxa. For examples *Akkermansia* and *Methanobrevibacter* were in higher relative abundance whereas *Prevotella* was in lower relative abundance[6,16,18]. However, there seems to be discrepancies in different study results with respect to experimental details and statistical analysis[19], in genetic and environmental dissimilarity[11] among MS patient cohorts or even disease treatment regimens[20].

A role of host genetics selecting and/or modulating gut microbiome in both healthy and diseased cohorts have been described particularly in type 1 diabetes and rheumatoid arthritis[21,22]. Genetic risk score enhances the predictive power of disease susceptibility and outcome[23]. A population with both genetic and environmental risk factors (GxE) are at a greater disease risk[24]. In this study, we showed that a cohort of MS patients have enhanced genetic risk score and also harbor a distinct gut microbiota which is different from the healthy controls suggesting an association between the genetic risk score and gut microbiota.

Methods
Study approval
This study obtained approval from institutional research board (IRB) of Marshfield Clinic Health System under IRB protocol SHU10417 and all the included subjects signed a written informed consent. The research was performed according to relevant guidelines and regulations. The reporting of this study followed most of the STORMS checklist for microbiome reporting studies[25].

Study design
Two-hundred thirty-seven MS patients and 50 controls were recruited in this case–control study from the Marshfield Clinic health Center (MCHS) during 2018–2021 who have had a recent diagnosis of MS (< 2 years of disease duration) or established diagnosis of MS (> 2 years of disease duration) regardless of clinical subtype (PPMS and RRMS) and treatment modality. The exclusion criteria were patients taking antibiotics, laxatives, or probiotics or who underwent a colonoscopy or similar procedure during the last three months.

All 237 patients provided a ~ 5.0 ml of blood samples while only 214 patients provided their fecal sample. All 50-control subjects provided both blood and fecal samples. We determined the 16S-based microbiota from 169 cases and 33 controls. The 169 patients were binned into five groups: treated RRMS (Group 1), treated PPMS (Group 2), treatment naïve RRMS but diagnosed for > 2 years of disease duration (Group 3), treatment naïve RRMS diagnosed for < 2 years of disease duration (Group 4), and treatment naïve PPMS (Group 5) as shown in Fig. 1. The patients included in groups one and two were on disease-modifying treatments (DMT) within six months of their stool collection. The DMTs were Glatiramer acetate, Dimethyl fumarate, Fingolimod, Natalizumab, Ocrelizumab or Teriflunomide. Groups two, four, and five were excluded from further analysis because each included < 10 patients, and 36 cases from Group 1 and Group 3 and seven healthy controls samples were filtered out due to low sequencing reads. The final microbiome analysis was based on 117 MS cases and 26 control subjects, and their demographic data is found in Table 1. Complete electronic health records (clinical and medications data) were not available for some non-MCHS patients who just sought MS clinical consultation/treatment at the MCHS.

Sample collection and storage
A self-collection fecal sample kit with detailed instructions was sent to each subject (patients and controls) or handed over by medical assistant of the caring physician to the MS patient during their routine visit with a provider. The fecal samples were returned in a boxed frozen cold pack to Dr. Shukla's laboratory where they were divided into aliquots and stored at −80 °C until further analysis. In addition, the recruited patients and controls provided a blood sample during their regular visit to MCHS's phlebotomy center. The blood samples were processed for serum, plasma, and buffy coat collection and stored at −80 °C.
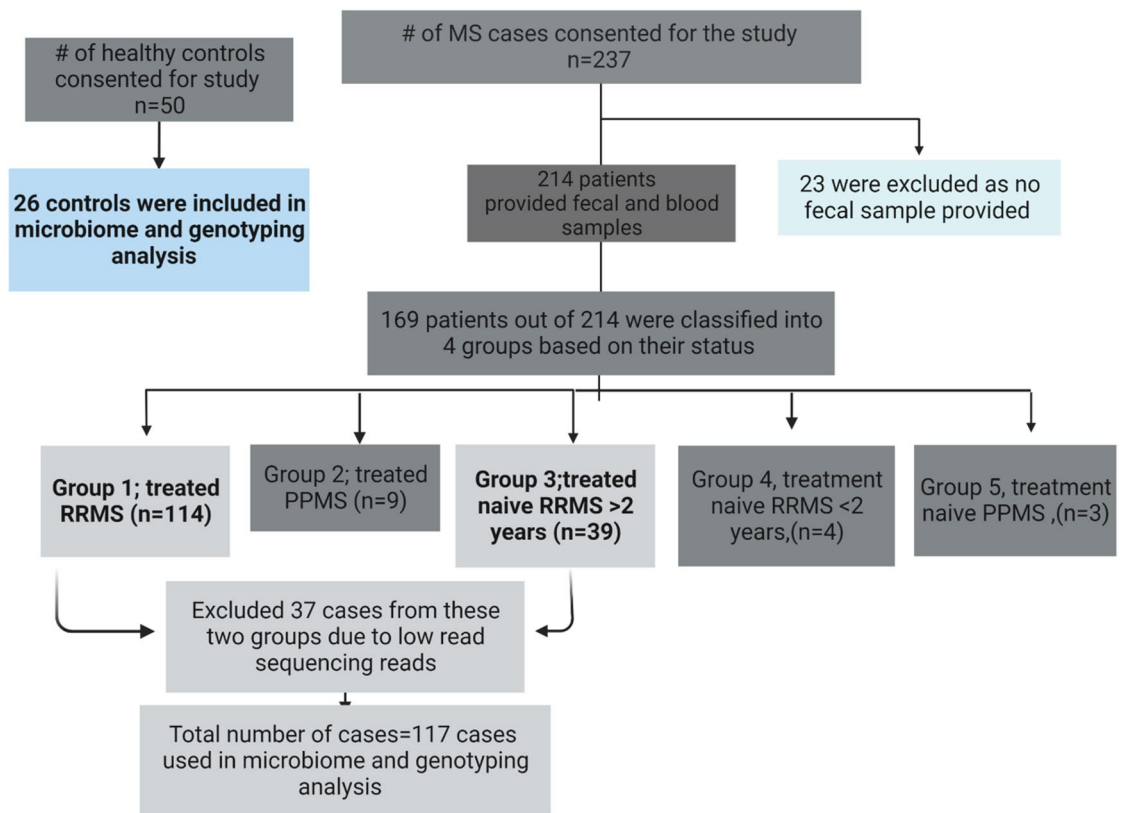
Figure 1. Flow chart displaying both the recruited MS patients and the healthy controls, their grouping and inclusion in the final analysis.

Characteristic	Group 1* (N = 83)	Group3** (N = 34)	Control (N = 26)
Age at diagnosis/enrollment in years (mean)	46.87	57.58	42.30
Sex (M/F)	(28/55)	(9/25)	(8/18)
Race	Caucasian	Caucasian	Caucasian##
BMI kg/m² (mean)	30.35	28.33	27.64
EBV flag***	5	0	0
Diabetes###	0	1	1
Hypertension###	18	12	3
Therapy			
Disease modifying therapy	76	0	0
Fingolimod (oral)	26		
Natalizumab (injectable)	6		
Ocrelizumab (injectable)	19		
Teriflunomide (oral pill)	6		
Glatrimer acetate	6		
Dimethyl fumarate	11		
Interferon beta-1a (injectable)	1		
Ritiximab	1		
Combination therapy#	6		

Table 1. Demographic characteristics of the MS patients and healthy control at the time of the stool collection. *Group 1 = treated RRMS patients. ** Group 3 = treatment naïve RRMS patients. ***EBV flag = Epstein-Barr virus previous infection. # On two disease modifying therapy within six months of fecal sample collection. ## All healthy control were white race except for one American Indian or Alaskan native. ### Incomplete electronic health record.

DNA extraction and 16S rRNA amplification

The microbial DNA was extracted from the fecal material using PowerLyzer PowerSoil DNA Isolation Kit (MoBio Laboratories, Inc., Carlsbad, CA) by following the manufacturer's protocol. Integrated DNA Technologies (Ames, IA) synthesized oligonucleotide primers (515F-806R) required for amplifying V4 region of 16S rRNA sequences[26] where the reverse amplification primer contained a 12 base barcode sequence and both primers contain adaptor regions[27]. The amplification was carried out using PE9700 thermocycler with the following run conditions initial denaturing temperature 94 °C for 2 min, 35 cycles of 94 °C for 45 s, 64 °C for 45 s and 72 °C for 45 s followed by a 10 min at 72 °C as final extension. SequalPrep™ Normalization Plate Kit was used to normalize the amplicon concentration (Thermofisher Scientific). Sequencing was done using the Illumina MiSeq Reagent Kit V2 with V4 sequencing primers as described by Caporaso et al. (2012)[28]. The total number of reads was 8,785,102 with an average read of 43,491.

Sequence data analysis

The demultiplexed paired-end reads from MiSeq were imported into Quantitative Insights Into Microbial Ecology (QIIME2, version 2019.10)[29] custom pipeline where the reads were assembled into one Fastq file identified with the sample names. Then, DADA2 plugin was used to denoise the sequences[30]. A fragment insertion tree using the q2-fragment-insertion plugin was created depending on alignment with the Greengenes database[31]. The generated Amplicon Sequence Variants (ASVs) from DADA2 were assigned to taxonomy using a pre-trained Naive Bayes classifier including the existing taxa in the 99% Greengenes 13_8 reference specific to the V4 hypervariable region corresponding to the primers we used[32]. A sampling depth of 24,520 reads was used to normalize the features count in each sample. The 16S microbiome analysis was performed on 117 cases and 26 controls.

Microbiome analysis

Alpha diversity was computed using the Faith's Phylogenetic Diversity (sum of the branch lengths of a phylogenetic tree connecting all species in the target assemblage[33]), Pielou's evenness[34] and Shannon indices using the Qiime 2 pipeline. Kruskal Wallis test was used to detect any significant differences between cases and controls in different indices of alpha diversity. Principal component analysis (PCA) based on unweighted unifrac[35] was carried out while doing the permutational multivariate analysis of variance (Permanova) test to detect if there was any significant difference between the clusters formed. Graphs were plotted using the ggplot2 package of the R statistical software 3.6.0. The relative abundance of a taxonomic unit for a grouping (e.g., Cluster A) was calculated by taking the average across each sample's relative proportion for that taxonomic unit and dividing by the total and multiplying by 100 to yield percentage. Taxonomic units that had a relative abundance of less than 1% were combined into a "Rare_combined" group. Welch's t-test was utilized to test for significant differences between groupings.

To detect a significant taxa at the phyla, family or genera level associated within the two clusters generated from PCA analysis, the Quasi-Conditional Association Test using General Estimating Equations (QCAT-GEE) was used, including a Permutation test[36]. The QCAT-GEE composes of three tests: the zero-test, which assess presence or absence of taxa, the positive-test, which assesses differences in abundance of each taxa, and the two-test, which combines the zero and positive-tests.

Genotyping and genetic risk score

Genotyping was performed on all 117 cases and 26 controls. Briefly, DNA from both the patients and healthy controls' buffy coat was isolated using QIAamp DNA blood mini kit (Qiagen Inc; Germanton, MD). The DNA samples were genotyped using Illumina's Infinium Immuno array-24 v2 chip at UW-Madison's Gene Expression Center (GEC). Variants were clustered and genotyped using GenomeStudio Data Analysis software 2.0 along with the chip manifest files. The SNPs were retained for imputation based on standard criteria (e.g., minimum allele frequency > 0.05; missingness < 0.01; individual genotype rate > 0.99; and Hardy–Weinberg equilibrium p-value > 1e−07)[37]. Genetic coverage was increased through imputation using genome build 38 Genotype Imputation HLA of the University of Michigan's Imputation Server[38]. A genetic risk score was calculated utilizing 187 relevant variants (Table 2 and Supplemental Table 1) previously identified by Patsopolous et al. (2019)[10]. The genetic risk score as defined by Chatterjee et al. (2016) is the quantitative measurement of the total genetic risk of multiple susceptibility variants (common, intermediate, and rare) of the disease[23]. The calculation of the genetic risk score for each subject was performed by summing the number of risk alleles for a given variant and multiplying the sum by the effect size obtained from Patsopolous et al., 2019. Plink software version 2.3.1 was then used to divide the score by the total number of SNPs[39].

Ethics approval and consent to participate

All included research subjects provided written consent and this research project was approved by the Institutional Review Board of Marshfield Clinic Health System (approval # IRB-19-447 and MCR Code: SHU10417).

Results

Demographics and summary of electronic health record from the study participants

The number of patients in Group 1 and 3 were 83 and 34 respectively. The average age of MS patients in Group 1 and Group 3 at diagnosis were 46.87 and 57.58 years respectively. Their average BMIs were 30.35 and 28.33 for Group 1 and Group 3, respectively. Seventy-six patients in Group 1 were on a single DMT while six patients were on two different DMTs in the last 6 months of the time of fecal samples collection (Table 1).

SNPs	OR	Locus in the MHC region
rs1071743	0.69	HLA-A
rs17493811	0.83	AGPAT1
rs3819292	1.09	HLA-B
AA B position 45 TK	1.13	HLA-B
rs4081559	1.31	HLA-B
rs3135024	1.16	DPA1/DPB1
rs3097671	1.34	DPB1
rs9277626	0.92	DPB2
rs11751659	1.17	DPB2
AA DQβ1 position-5 L	1.24	DQB1
rs766848979 A	0.84	DRB1
rs67476479 CA	1.32	DRB1
HLA-DRB1*01:03	2.9	DRB1
rs9271366	1.57	intergenic (DRB1/DQA1)
rs114071505	0.78	Intergenic (RNF39/TRIM31)
rs9266629	0.82	intergenic (ZDHHC20P2/FGFR3P1)
rs2844482	1.35	LST1 (class III haplotype)
rs2229092	1.17	LTA
rs3093982	1.11	MCCD1
rs2523500	0.92	NFKBIL1

Table 2. Studied SNPs associated with MS inside MHC region. *LST1* leukocyte specific transcript 1, *LTA* lymphotoxin-α, *MCCD1* mitochondrial coiled-coil domain 1, *NFKBIL1* NF-κB inhibitor-like protein 1.

The gut microbiome profile of Group 1 (Treated RRMS) and Group 3 (Treatment naïve RRMS) MS cases and controls

The Faith's phylogenetic diversity between the MS cases and non-MS healthy controls was significantly different (Fig. 2A, p = 0.002) and so was the Pielou's evenness index (Fig. 2B, p = 0.03). However, the Shannon diversity index between the cases and controls was not significantly different (Fig. 2C). When we compared the Faith's phylogenetic diversity between Group 1, Group 3, and healthy control group (see materials and methods), we observed that while both Group 1 and Group 3 were significantly different from the healthy control group, the difference was not significant between the two case groups (Fig. 3).

Identification of a unique MS patients' cluster

When we performed the unweighted UniFrac principal component analysis (PCA) on microbiota of 117 cases and 26 controls, we observed two clusters, a large cluster named Cluster A consisting of 98 cases and 26 control (n = 124) and a smaller cluster named Cluster B consisting of 19 cases only (Fig. 4). PC1 accounted for 15.1% of the variation, while PC2 accounted for 9.94% of the variation. These two clusters were significantly different by the Permanova test (*p* = 0.01). However, differences in these two clusters were not associated with age, DMT used, number of MRI lesions or any other disease conditions like gastric issues.

Relative abundance analysis

We observed several differences in relative abundances of different phyla at 99% cutoff, but they were not statistically significant. For example, *Actinobacteria* showed a higher relative abundance in Cluster B whereas *Verrucomicrobia* showed a higher relative abundance in cluster A (Fig. 5A, and Supplemental Table 2). Both *Bacteroidetes* and *Firmicutes* showed comparable abundances in the two clusters. Interestingly, *Proteobacteria* was not detected in cluster A. As shown in Fig. 5B and Supplemental Table 3, *Lachnospiraceae* family showed small difference between the two clusters. At the genus levels, Cluster B lacks *Phascolarctobacterium* while cluster A lacks both *Clostridium* and *Megasphaera* genera (Fig. 5C and Table 3). Moreover, *Bifidobacterium* showed higher abundance in Cluster B while *Akkermansia* in cluster A.

The QCAT-GEE tests showed difference between cluster A and cluster B

After applying all three tests of QCAT-GEE to the taxonomy table with all ranks from kingdom to genus, we observed that QCAT-GEE two-parts test and the positive test showed that only *Porphyromonadaceae family* was significantly different between two clusters A and B (*p* = 0.008).

Genetic risk scores associated with clusters A and B

Since MS has a strong genetic susceptibility component, we used a validated risk score method with MS disease[10,40] in our study. The genetic background of the patients were self-identified as Caucasian. Population stratification was examined by PCA where eight individuals were determined to be outliers and were removed
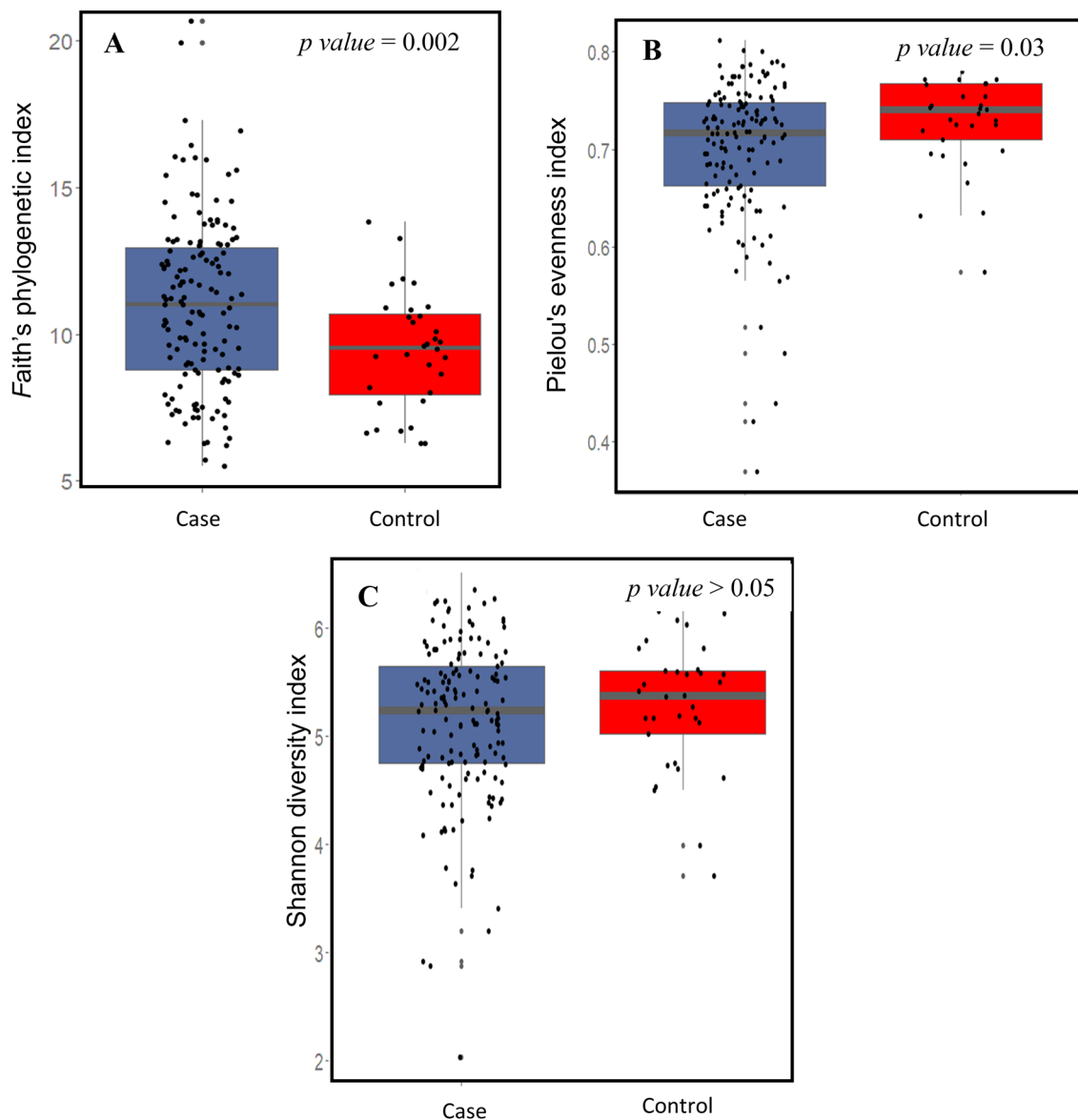
Figure 2. (**A**) The boxplot representing (**A**) Faith's Phylogenetic diversity (PD) where there was a significant difference between MS cases and control. (**B**) Pielou's evenness index where there was a significant difference between cases and controls. (**C**) Shannon–wiener diversity index (H) where both MS cases and control microbiome were similar, Kruskal Wallis test was used to detect any significant differences between cases and controls in different indices of alpha diversity.

from further analysis. The analysis was repeated and it was determined that the impact on the results of these removals did not change the interpretations or conclusions. The healthy control subjects tended to have a lower genetic risk score (from 0.007 to 0.017) whereas, the MS cases tended to have a higher genetic risk score from 0.007 to 0.022 (Fig. 6A). The t-test also showed high significant differences between the genetic risk scores between cases and controls ($p = 2.682e{-}05$). When considering both the microbiome diversity-based clusters and the genetic risk scores together (Fig. 6B), an interesting trend was observed where the gut-microbiome of subjects of Cluster A, which included a significant number of both cases and controls, tended to have a lower genetic risk scores compared to cluster B (higher genetic risk score) which consisted of cases only. The difference in genetic score between controls and two cases groups was statistically significant by t-test. This suggests that the patients with higher genetic risk score may be associated with a unique gut microbiome. Additionally, the cases with lower genetic risk scores tend to have their microbiome closer to the healthy controls compared to the cases with a higher genetic risk score.

Discussion

Unravelling the genetic-environmental factor(s) that control susceptibility to complex autoimmune diseases such as MS is challenging. Studies have identified the genetic susceptibilities to MS[10,12,13] and association of individual microbes and/or gut dysbiosis in MS[5,6,16]. In this study, we calculated a genetic risk score of MS patients in our
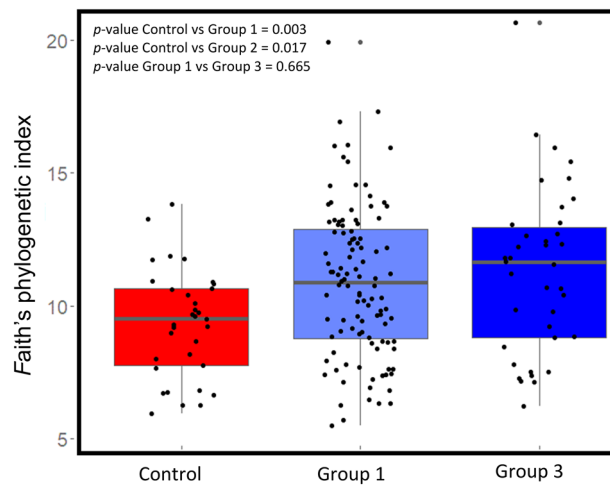
Figure 3. Box plot representing the Faith's phylogenetic diversity in MS patients group 1 (Group 1, treated RRMS), group 3 (Group 3, treatment naïve RRMS), and healthy controls. There was a significant difference between the control and each group of MS cases individually. On the other side, there was no significant difference between the two groups of MS cases.
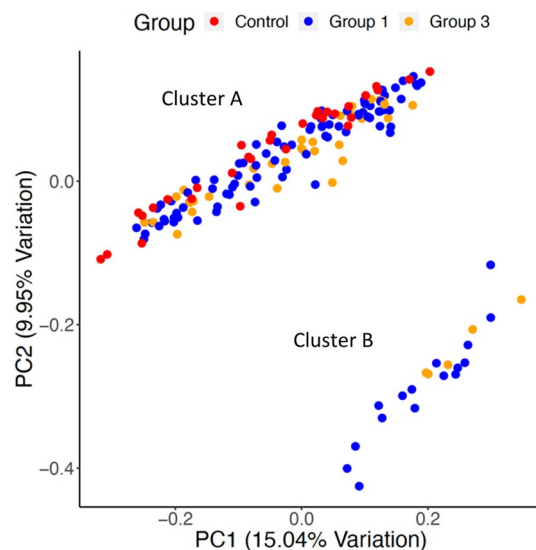

Figure 4. Unweighted UniFrac Principal Coordinate (PCoA) of groups one (treated RRMS), three (treatment naïve RRMS) and healthy controls showed two distinct clusters (A and B). Each dot represents a MS case or healthy control and the PCoA plot show the abundant taxa in each patient gut microbiota. The two chosen PC coordinates showed the most diversity and the diversity captured is represented in percentage on the axis.

cohorts and conducted gut microbiome analyses to see the association between high-risk cohorts and identify unique microbiota signatures in their gut. We calculated the genetic risk score based on a validated risk scores for the MS disease as described before[10]. Since MS is a complex disease, many variants share the responsibility in increasing the patients' susceptibility to this devastating disease. Surprisingly, we found that patients exhibiting the highest genetic risk score are the patients who had a distinct microbiome.

Through studying the demographics data of the patients enrolled in this study and their Epstein-Barr Virus (EBV) status in the electronic health record, we made several observations. Notably, only five out of the tested MS patients had evidence of EBV infection in their electronic health records (see Table 1). This was surprising as EBV was recently described as a virus that increases the risk for MS susceptibility[41]. As expected, the female ratio was higher than males which is known in MS disease epidemiology[42].

We observed a significant difference in both Faith's phylogenetic diversity and Pielou's evenness indices between the MS cases and healthy controls. The higher significant faith phylogenetic diversity MS cases was in agreement with the unweighted UniFrac analysis. Absence of any significant differences in Shannon index between our MS case and healthy control cohorts was similar to a couple of previously published results,
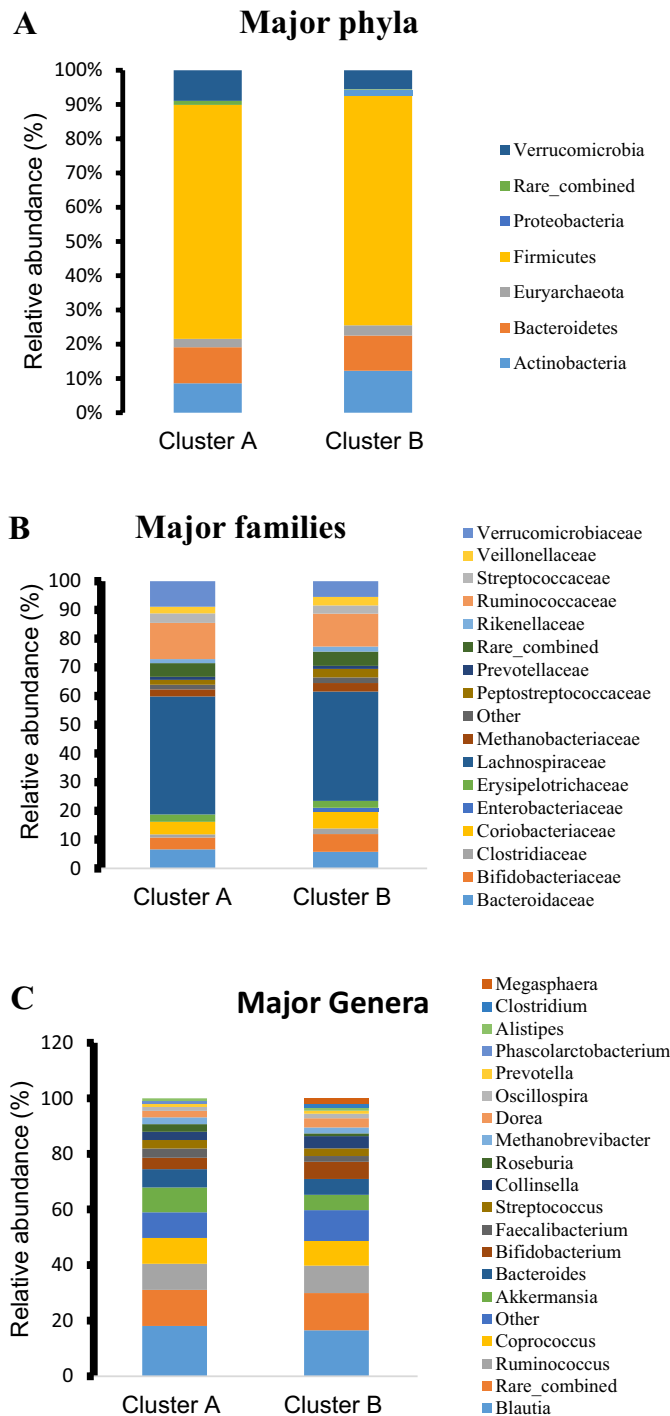
Figure 5. Relative abundance of major phyla (**A**), major families (**B**), and major genera (**C**) in the clusters A and B generated from the Unweighed Unifrac.

especially if the gut microbiome was collected during non-active episodes of the MS disease[4,43]. Indeed, Chen et al. (2016) reported that patients with active episodes of RRMS have a decline in the species richness[16]. There were no significant differences in alpha and beta diversity indices in patients treated with DMTs when compared with treatment naïve patients. Some studies have reported changes in gut microbiome composition after treatment especially with glatiramer acetate (GA) and dimethyl fumarate (DMF)[20,44]. Future studies should consider assessing the gut microbiome of MS patients at different time points in RRMS to ascertain different dysbiosis state during the active and non-active phases of the disease. Surprisingly, we observed a significant number of MS patients who had a higher genetic risk score than the healthy controls but had a similar gut microbiome compared to the controls. We speculate that this could be due to fact that (i) 94% (91 out of 98) of the patients

Genus	Cluster A (%)	Cluster B (%)
Blautia	18.06	16.50
Coprococcus	9.28	8.89
Akkermansia	8.92	5.51
Bacteroides	6.58	5.73
Ruminococcus	9.38	9.82
Bifidobacterium	4.10	6.17
Collinsella	2.95	4.41
Streptococcus	3.07	2.72
Roseburia	2.75	1.02
Faecalibacterium	3.33	2.06
Methanobrevibacter	2.41	2.12
Dorea	2.33	3.28
Prevotella	1.07	1.08
Alistipes	1.00	1.09
Clostridium	0	1.52
Oscillospira	1.41	1.61
Phascolarctobacterium	1.05	0
Megasphaera	0	1.52

Table 3. Comparative relative genera abundance between cluster A and cluster B.

were in remitting phase of which 68% (67 out of 98) were also on one or more DMTs. The unweighted UniFrac analysis is sensitive to rare lineages within a microbial community[45,46]. Based on our QCAT analysis, *Porphyromonadaceae*, a typically low abundance family, was observed to be significantly higher in Cluster B. This family was also found to be associated with worse EAE outcomes in genetically susceptible mice[47]. Moreover, its presence has been correlated with IL17, an interleukin known to be upregulated in MS which is also found to be high in ileum in other metabolic diseases like obesity and diabetes[48]. Even in systemic inflammatory diseases such as ankylosing arthritis[49] and neurodegenerative disease such as Parkinson disease, the family *Porphyromonadaceae* was enriched[50]. In our study, lower number of patient samples in Cluster B limited the identification of other significant taxa associated with high genetic risk score patients. Meanwhile, *Akkermansia* showed non-significant higher abundance in Cluster A as reported before[51].

Genetic studies conducted to detect the variants causing MS disease reported the modest effect to variant HLA DRB1*15:01 and many other loci with smaller associations[52]. However, no study can point the risk of MS disease to a certain allele in the HLA class II because all of the alleles in this region, especially, in European ancestors are inherited together due to intense linkage disequilibrium[54]. Moreover, some of these detected SNPs for MS risk in the literature are common and can be present in healthy unaffected individuals[53]. Thus, genetic risk score measurement is suitable in this complex disease to reveal the cumulative risk to MS. Good predictability was achieved by measuring this score before in other diseases (prostate cancer and systemic lupus erythematosus) including MS disease[55–57]. In our cohort, as expected the MS patients have higher genetic risk score than healthy controls. However, we found a unique cohort with the highest risk score having a unique gut microbiome. It is not clear at this stage, whether the dysbiotic gut microbiome increased the MS risk together with the genetic susceptibility or the host genotype affected the gut microbiome composition. For instance some studies have suggested the heritability of the gut microbiome[5]. Studies have also suggested that host genes affect the shape of the gut habitat thereby leading to variation in the gut microbiome[58]. Furthermore, in case of the MS, variants in MHC region in general could affect the shaping the gut microbiome through restricted colonization of some bacterial species through either their immune elimination or their inability to adhere to the intestinal epithelium[59] through the IgA mediated selection[60]. In addition, MHC region affects the T-cells maturation which subsequently can affect its autoreactivity[61]. Indeed, all of these studies support the interaction of genes and gut microbiome in precipitating different diseases.

Conclusion

In summary, we showed that a small cohort of MS patients showed high genetic risk score who also harbored a distinct microbiota in the gut. This observation drove the idea that indeed, genetic susceptibility in combination with the dysbiosis of the gut microbiome is associated with MS albeit in a smaller number of patients in our study. While future studies with larger cohorts of patients are needed to confirm the relationship between the genetic risk score and the MS gut microbiome, we believe our study provides a foundation for such a study.
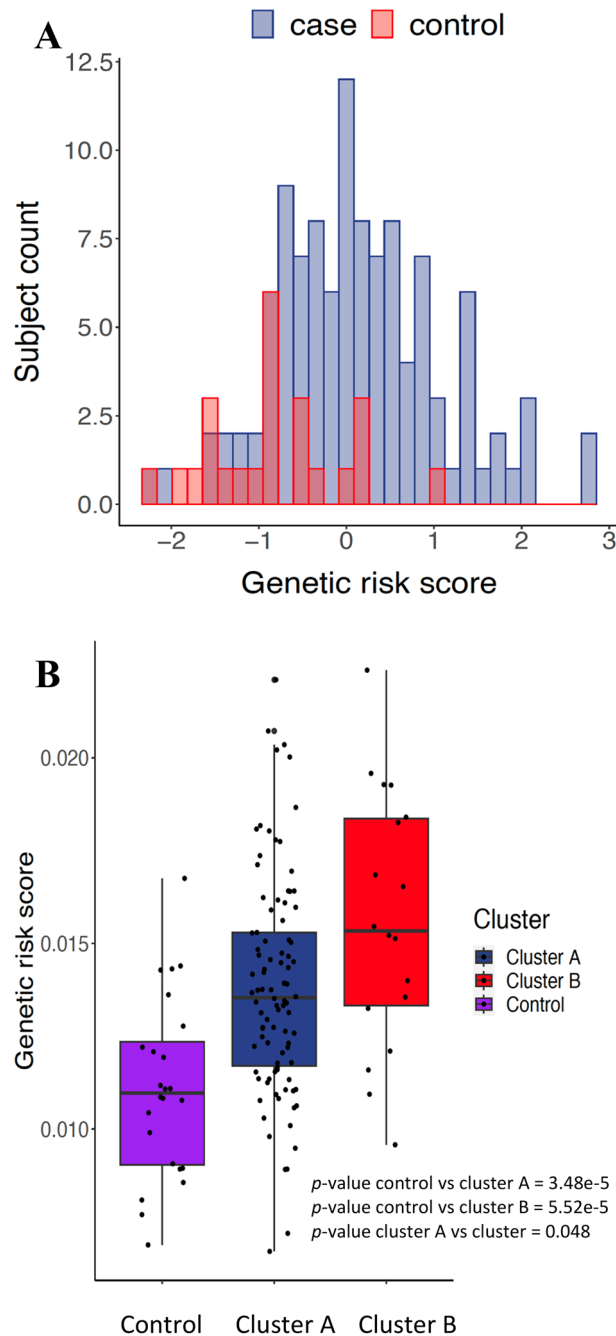
Figure 6. (**A**) A Histogram showing the distribution of genetic risk score in our cohort of both healthy controls (red bars) and MS cases (blue bars). Here, the histogram showed the controls having low genetic risk scores in comparison to the MS cases. The genetic risk scores were tested using SNPs both inside MHC region (Table 2) and outside MHC region (supplementary Table 1). T test showed high significant difference between cases and control ($p = 2.682e-05$). (**B**) The genetic risk score of the subjects included in the two generated clusters (A and B) and the healthy controls in cluster A from the Unweighted unifrac analysis. This plot showed the rising of genetic risk scores from low values in healthy controls to higher values in the both clusters comprising of MS cases. However, the cluster B showed the highest genetic risk score. T test to check difference in means between control and cluster A ($p = 3.48e-5$) and control and cluster B ($p = 5.52e-5$) showed significant difference. T test showed significant difference between cluster A and B ($p = 0.048$).

Data availability

All sequence files and metadata for all samples used in this study have been deposited in NCBI under BioProject number PRJNA889427. A full record of all statistical analysis is included as supplementary files 1 and 2.

References

1. Walton, C. *et al.* Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition. *Mult. Scler. (Houndmills Basingstoke Engl. 2020/11/11 ed. SAGE Publications)* **26**, 1816–1821 (2020).
2. Kozhieva, M. *et al.* Primary progressive multiple sclerosis in a Russian cohort: Relationship with gut bacterial diversity. *BMC Microbiol. BioMed. Central* **19**, 309–309 (2019).
3. Compston, A. & Coles, A. Multiple sclerosis. *Lancet* **372**, 1502–1517 (2008).
4. Cosorich, I. *et al.* High frequency of intestinal T(H)17 cells correlates with microbiota alterations and disease activity in multiple sclerosis. *Sci. Adv. (American Association for the Advancement of Science)* **3**, e1700492–e1700492 (2017).
5. Berer, K. *et al.* Gut microbiota from multiple sclerosis patients enables spontaneous autoimmune encephalomyelitis in mice. *Proc. Natl. Acad. Sci. USA (2017/09/11 ed. National Academy of Sciences)* **114**, 10719–10724 (2017).
6. Jangi, S. *et al.* Alterations of the human gut microbiome in multiple sclerosis. *Nat. Commun. (Nature Publishing Group)* **7**, 12015–12015 (2016).
7. Holmøy, T., Kvale, E. Ø. & Vartdal, F. Cerebrospinal fluid CD4+ T cells from a multiple sclerosis patient cross-recognize Epstein-Barr virus and myelin basic protein. *J. Neurovirol.* **10**, 278–283 (2004).
8. Zostawa, J., Adamczyk, J., Sowa, P. & Adamczyk-Sowa, M. The influence of sodium on pathophysiology of multiple sclerosis. *Neurol. Sci. (2017/01/11 ed. Springer Milan)* **38**, 389–398 (2017).
9. Dopkins, N., Nagarkatti, P. S. & Nagarkatti, M. The role of gut microbiome and associated metabolome in the regulation of neuroinflammation in multiple sclerosis and its implications in attenuating chronic inflammation in other inflammatory and autoimmune disorders. *Immunology (2018/02/27 ed. John Wiley and Sons Inc.)* **154**, 178–185 (2018).
10. International Multiple Sclerosis Genetics Consortium. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science* **365**, eaav7188 (2019).
11. Montgomery, T. L. *et al.* Interactions between host genetics and gut microbiota determine susceptibility to CNS autoimmunity. *Proc. Natl. Acad. Sci.* **117**, 27516 (2020).
12. Goris, A., Vandebergh, M., McCauley, J. L., Saarela, J. & Cotsapas, C. Genetics of multiple sclerosis: Lessons from polygenicity. *Lancet Neurol. Elsevier* **21**, 830–842 (2022).
13. Moutsianas, L. *et al.* Class II HLA interactions modulate genetic risk for multiple sclerosis. *Nat. Genet.* **47**, 1107–1113 (2015).
14. Hollenbach, J. A. & Oksenberg, J. R. The immunogenetics of multiple sclerosis: A comprehensive review. *J. Autoimmun.* **64**, 13–25 (2015).
15. Patsopoulos, N. A. & De Jager, P. L. Genetic and gene expression signatures in multiple sclerosis. *Mult. Scler. J. (SAGE Publications Ltd STM)* **26**, 576–581 (2020).
16. Chen, J. *et al.* Multiple sclerosis patients have a distinct gut microbiota compared to healthy controls. *Sci. Rep. (Nature Publishing Group)* **6**, 28484–28484 (2016).
17. Berer, K. *et al.* Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination. *Nature* **479**, 538–541 (2011).
18. Elsayed, N. S., Aston, P., Bayanagari, V. R. & Shukla, S. K. The gut microbiome molecular mimicry piece in the multiple sclerosis puzzle. *Front. Immunol.* https://doi.org/10.3389/fimmu.2022.972160 (2022).
19. Wallen, Z. D. Comparison study of differential abundance testing methods using two large Parkinson disease gut microbiome datasets derived from 16S amplicon sequencing. *BMC Bioinform. BioMed. Central* **22**, 265–265 (2021).
20. Katz Sand, I. *et al.* Disease-modifying therapies alter gut microbial composition in MS. *Neurol. Neuroimmunol. Neuroinflamm. (Lippincott Williams & Wilkins)* **6**, e517–e517 (2018).
21. Russell, J. T. *et al.* Genetic risk for autoimmunity is associated with distinct changes in the human gut microbiome. *Nat. Commun. (Nature Publishing Group UK)* **10**, 3621–3621 (2019).
22. Wells, P. M. *et al.* Associations between gut microbiota and genetic risk for rheumatoid arthritis in the absence of disease: a cross-sectional study. *Lancet Rheumatol. (Elsevier)* **2**, e418–e427 (2020).
23. Chatterjee, N., Shi, J. & García-Closas, M. Developing and evaluating risk prediction models for stratified disease prevention. *Nat. Rev. Genet.* **17**, 392–406 (2016).
24. Lewis, C. M. & Vassos, E. Genetic risk scores: From research tools to clinical instruments. *Genome Med. BioMed. Central* **12**, 44–44 (2020).
25. Mirzayi, C. *et al.* Reporting guidelines for human microbiome research: The STORMS checklist. *Nat. Med.* **27**, 1885–1892 (2021).
26. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
27. Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci.* **108**, 4516–4522 (2011).
28. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* **6**, 1621–1624 (2012).
29. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
30. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
31. Janssen, S. *et al.* Phylogenetic placement of exact amplicon sequences improves associations with clinical information. *mSystems (American Society for Microbiology)* **3**, e00021-e118 (2018).
32. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
33. Faith, D. P. Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* **61**, 1–10 (1992).
34. Pielou, E. C. The measurement of diversity in different types of biological collections. *J. Theor. Biol.* **13**, 131–144 (1966).
35. Lozupone, C. A., Hamady, M., Kelley, S. T. & Knight, R. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl. Environ. Microbiol. (American Society for Microbiology)* **73**, 1576–1585 (2007).
36. Tang, Z.-Z., Chen, G., Alekseyenko, A. V. & Li, H. A general framework for association analysis of microbial communities on a taxonomic tree. *Bioinform. Oxf. Engl. (Oxford University Press)* **33**, 1278–1285 (2017).
37. Shi, S. *et al.* Comprehensive assessment of genotype imputation performance. *Hum. Hered.* **83**, 107–116 (2018).
38. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
39. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
40. Cortes, A. & Brown, M. A. Promise and pitfalls of the Immunochip. *Arthritis Res. Ther.* **13**, 101 (2011).
41. Kjetil, B. *et al.* Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis. *Science (American Association for the Advancement of Science)* **375**, 296–301 (2022).
42. Harbo, H. F., Gold, R. & Tintoré, M. Sex and gender issues in multiple sclerosis. *Ther. Adv. Neurol. Disord. (SAGE Publications)* **6**, 237–248 (2013).

43. Miyake, S. *et al.* Dysbiosis in the gut microbiota of patients with multiple sclerosis, with a striking depletion of species belonging to *Clostridia* XIVa and IV clusters. *PloS One (Public Library of Science)* **10**, e0137429–e0137429 (2015).
44. Boziki, M. K. *et al.* Microbiome in multiple sclerosis: Where are we, what we know and do not know. *Brain Sci.* **10**, 234 (2020).
45. Chen, J. *et al.* Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics* **28**, 2106–2113 (2012).
46. Allen, J. M. *et al.* Voluntary and forced exercise differentially alters the gut microbiome in C57BL/6J mice. *J Appl Physiol. (American Physiological Society)* **118**, 1059–1066 (2015).
47. Cox, L. M. & Weiner, H. L. The microbiome requires a genetically susceptible host to induce central nervous system autoimmunity. *Proc. Natl. Acad. Sci.* **117**, 27764–27766 (2020).
48. Garidou, L. *et al.* The gut microbiota regulates intestinal CD4 T cells expressing RORγt and controls metabolic disease. *Cell Metab.* **22**, 100–112 (2015).
49. Forbes, J. D., Van Domselaar, G. & Bernstein, C. N. The gut microbiota in immune-mediated inflammatory diseases. *Front. Microbiol. (Frontiers Media S.A.)* **7**, 1081–1081 (2016).
50. Romano, S. *et al.* Meta-analysis of the Parkinson's disease gut microbiome suggests alterations linked to intestinal inflammation. *Npj Park Dis.* **7**, 27 (2021).
51. Mirza, A. *et al.* The multiple sclerosis gut microbiota: A systematic review. *Mult. Scler. Relat. Disord.* **37**, 101427 (2020).
52. Patsopoulos, N. A. Genetics of multiple sclerosis: An overview and new directions. *Cold Spring Harb. Perspect. Med. (Cold Spring Harbor Laboratory Press)* **8**, 8951 (2018).
53. Gresle, M. M. *et al.* Multiple sclerosis risk variants regulate gene expression in innate and adaptive immune cells. *Life Sci. Alliance* **3**, e202000650 (2020).
54. Lincoln, M. R. *et al.* Epistasis among HLA-DRB1, HLA-DQA1, and HLA-DQB1 loci determines multiple sclerosis susceptibility. *Proc. Natl. Acad. Sci. USA (National Academy of Sciences)* **106**, 7542–7547 (2009).
55. Helfand, B. T. A comparison of genetic risk score with family history for estimating prostate cancer risk. *Asian J Androl. (Medknow Publications & Media Pvt Ltd)* **18**, 515–519 (2016).
56. Reid, S. *et al.* High genetic risk score is associated with early disease onset, damage accrual and decreased survival in systemic lupus erythematosus. *Ann. Rheum. Dis. (BMJ Publishing Group)* **79**, 363–369 (2020).
57. De Jager, P. L. *et al.* Integration of genetic risk factors into a clinical algorithm for multiple sclerosis susceptibility: A weighted genetic risk score. *Lancet Neurol.* **8**, 1111–1119 (2009).
58. Spor, A., Koren, O. & Ley, R. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat. Rev. Microbiol.* **9**, 279–290 (2011).
59. Toivanen, P., Vaahtovuo, J. & Eerola, E. Influence of major histocompatibility complex on bacterial composition of fecal flora. *Infect. Immunol. (American Society for Microbiology)* **69**, 2372–2377 (2001).
60. Kubinak, J. L. *et al.* MHC variation sculpts individualized microbial communities that control susceptibility to enteric infection. *Nat. Commun. (Nature Pub. Group)* **6**, 8642–8642 (2015).
61. Shahi, S. K., Ali, S., Jaime, C. M., Guseva, N. V. & Mangalam, A. K. HLA class II polymorphisms modulate gut microbiota and experimental autoimmune encephalomyelitis phenotype. *ImmunoHorizons.* **5**, 627–646 (2021).

Acknowledgements

Author contributions

S.K.S., P.A., K.A., and R.V. conceived and designed the study, T.K. along with S.K.S. and P.A. enrolled the study participants, N.E. and T.L. performed the laboratory work, S.K.S., V.G., N.E., R.V., V.B., Q.L., J.M., Z.T., T.L., and T.K. analyzed the data, N.E., S.K.S., R.V., wrote the first draft and all other authors subsequently contributed to the writing and editing of the manuscript.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-43217-4.

Correspondence and requests for materials should be addressed to S.K.S.

Reprints and permissions information is available at www.nature.com/reprints.

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