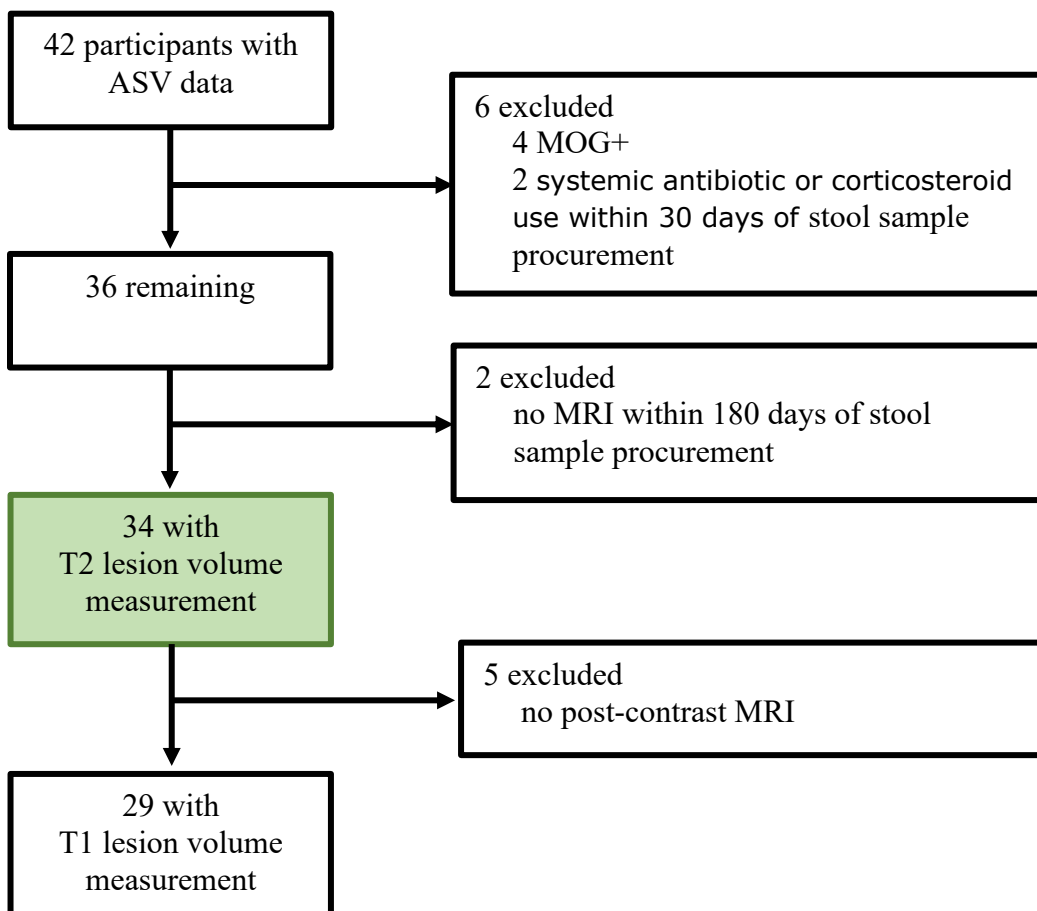


Supplementary Materials

A Cross-Sectional Study of MRI Features and the Gut Microbiome in Pediatric-Onset Multiple Sclerosis

Feng Zhu, Yinshan Zhao, Douglas L Arnold, Amit Bar-Or, Charles N Bernstein, Christine Bonner, Morag Graham, Janace Hart, Natalie Knox, Ruth Ann Marrie, Ali I Mirza, Julia O'Mahony, Gary Van Domselaar, E. Ann Yeh, Brenda Banwell, Emmanuelle Waubant, Helen Tremlett
US Network of Pediatric MS Centers, the Canadian Pediatric Demyelinating Disease Network

Figure S1 Cohort flow chart



Abbreviations:

ASV = amplicon sequence variants

MOG+ = positive for serum myelin oligodendrocyte glycoprotein antibodies

(29 had a MOG test and 13 did not have a test)

Alpha Diversity

In the microbiome context, alpha diversity describes the diversity within a microbial community. Different alpha diversity measures exist which summarize the richness (total number of taxonomic groups present in the community), evenness (relative abundance of the groups), or both.

Table S1 Alpha diversity measures used in this study

Diversity index	Shannon index
Description	A measure that reflects both the evenness and the richness. The Shannon index quantifies the uncertainty in predicting the identity of a taxonomic member that is randomly selected from the community. The index value increases as the number of taxonomic groups increases and as the groups distribute more evenly.
Formula	$-\sum_{i=0}^s (p_i \ln p_i)$
Reference	Shannon, C.E. <i>A mathematical theory of communication</i> . Bell Syst. Tech. J, 1948. 27 : p. 379–423.
Diversity index	Margalef's index
Description	A measures richness adjusted for library size (the total number of sequences in the sample).
Formula	$\frac{s}{\ln N}$
Reference	Margalef, R. <i>Information theory in ecology</i> . General Systems, 1958. 3 : p. 36–71.
Diversity index	Chao1 index
Description	A measure of richness which is an estimate of the expected number of taxonomic groups in the community based on the observed number of groups in the sample. The Chao1 index accounts for unobserved rare taxa and gives more weight to the low abundance groups.
Formula	$s + \frac{F_1(F_1 - 1)}{2(F_2 + 1)}$
Reference	Chao A. <i>Non-parametric estimation of the number of classes in a population</i> . Scand. J. Stat, 1984. 11 : p. 265-270.

Notation: s is the total number of taxonomic groups in the sample, p_i is the proportion of the sample represented by group i , N is the total number of sequences in the sample (library size), and, F_1 and F_2 are the number of groups that are only observed once or twice in the sample.

Table S2 Pearson’s product-moment correlation between participant characteristics and lesion volumes

Characteristic	T2 lesion volume Correlation (p-value)	T1 lesion volume Correlation (p-value)
Age at MRI	0.33 (0.06)	0.33 (0.08)
Age at symptom onset	0.18 (0.30)	0.15 (0.43)
Disease duration at MRI	0.24 (0.17)	0.25 (0.20)

Table S3 P-value for comparison of lesion volumes by participant characteristics based on the Kruskal–Wallis rank sum test

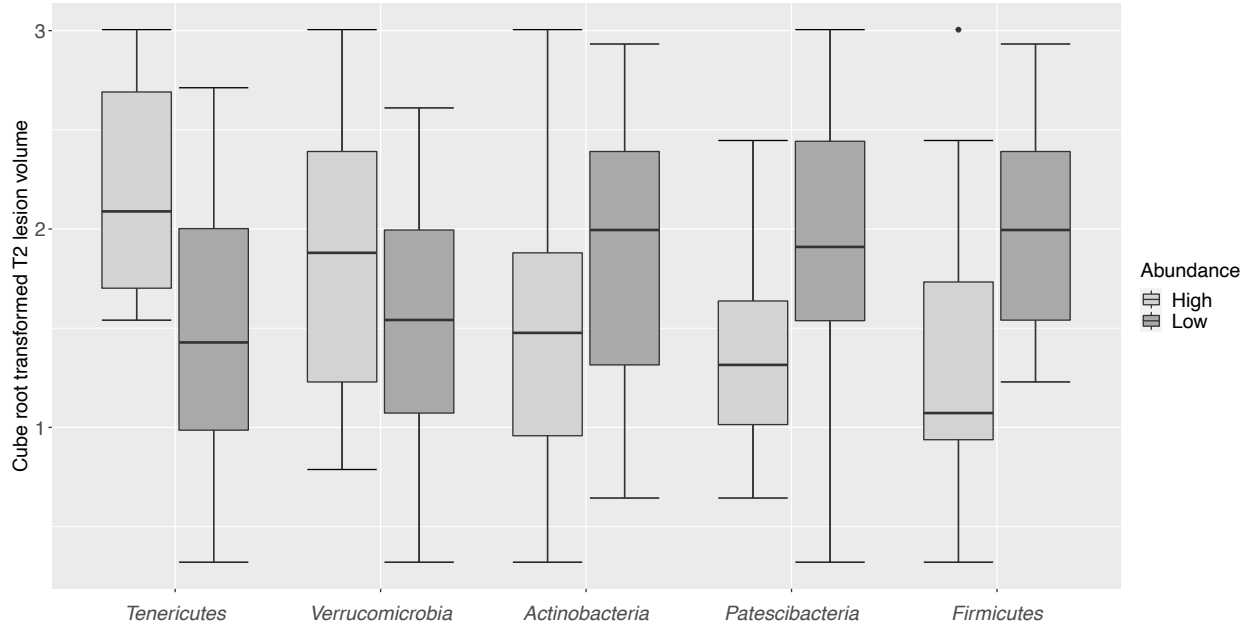
Characteristic	T2 lesion volume	T1 lesion volume
Sex	0.17	0.17
Bristol Stool Scale (hard, normal, loose)	0.68	1.00
Country of residence (Canada vs USA)	0.71	0.59
Body Mass Index (obese vs normal)	0.30	0.11
Disease modifying drug exposure status (ever vs never)	0.19	0.16

Table S4 Number of participants with non-zero count and median relative abundance of taxa selected by the Lasso Regression

Phylum			Genus		
Name	Non-Zero Counts (n)	Relative Abundance (Median)	Name	Non-Zero Counts (n)	Relative Abundance (Median)
<i>Tenericutes</i>	8	0	-	-	-
<i>Verrucomicrobia</i>	20	57	-	-	-
<i>Actinobacteria</i>	34	5942	<i>Atopobiaceae; Uncultured Adlercreutzia</i>	7 23	0 56
<i>Patesibacteria</i>	15	0	-	-	-
<i>Firmicutes</i>	34	54967	<i>Ruminiclostridium</i>	10	0
			<i>Ruminococcaceae NK4A214</i>	23	55
			<i>Butyricoccus</i>	32	93
			<i>Coprococcus 3</i>	29	377
			<i>Eubacterium eligens</i>	24	65
			<i>Moryella</i>	19	5
			<i>Roseburia</i>	31	158
			<i>Ruminococcus gnavus</i>	20	3
			<i>Intestinibacter</i>	32	119
			<i>Candidatus Stoquefichus</i>	7	0
			<i>Erysipelatoclostridium</i>	27	26

Figure S2 Boxplots of T2 and T1 lesion volumes by abundance level of the phyla selected from the Lasso regression

a) T2 lesion volume



b) T1 lesion volume

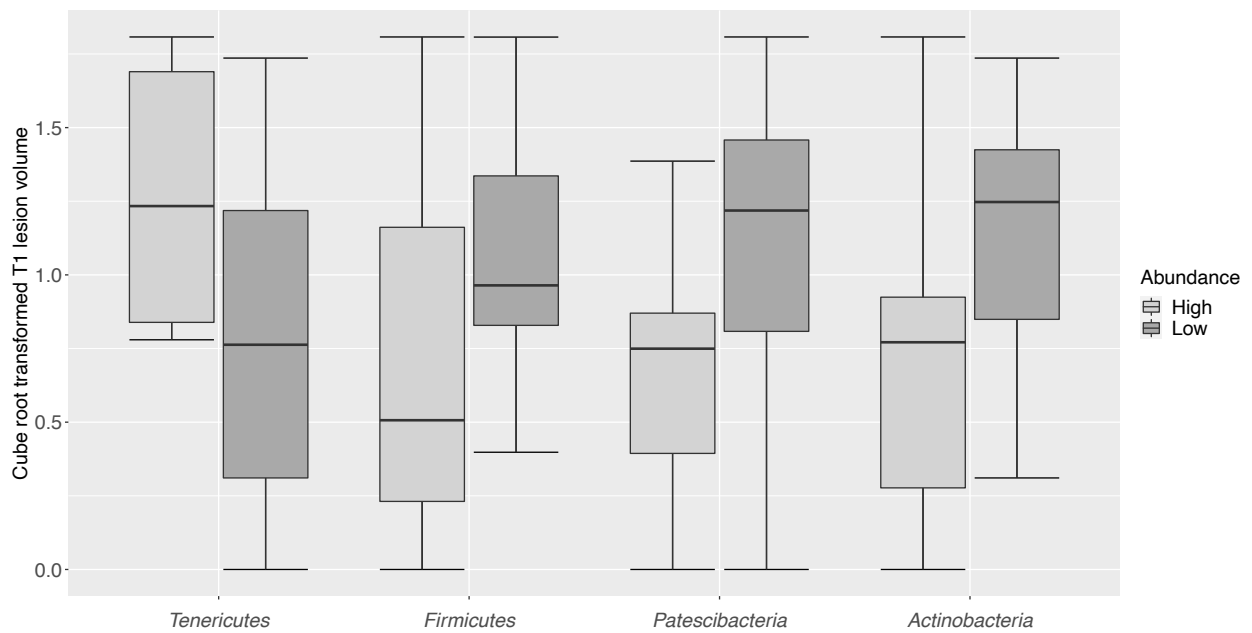


Figure S3 Lineage of the genera selected from the Lasso regression

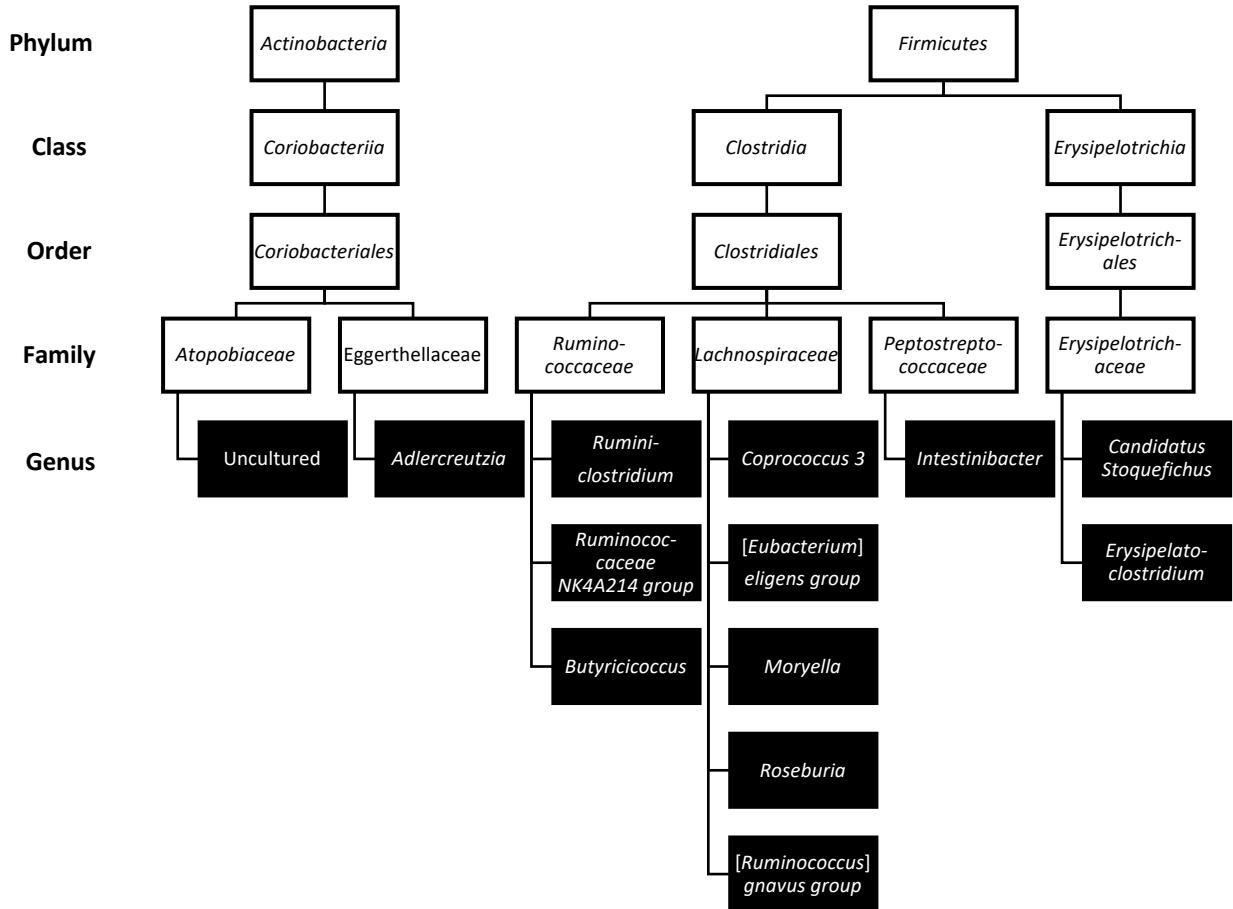
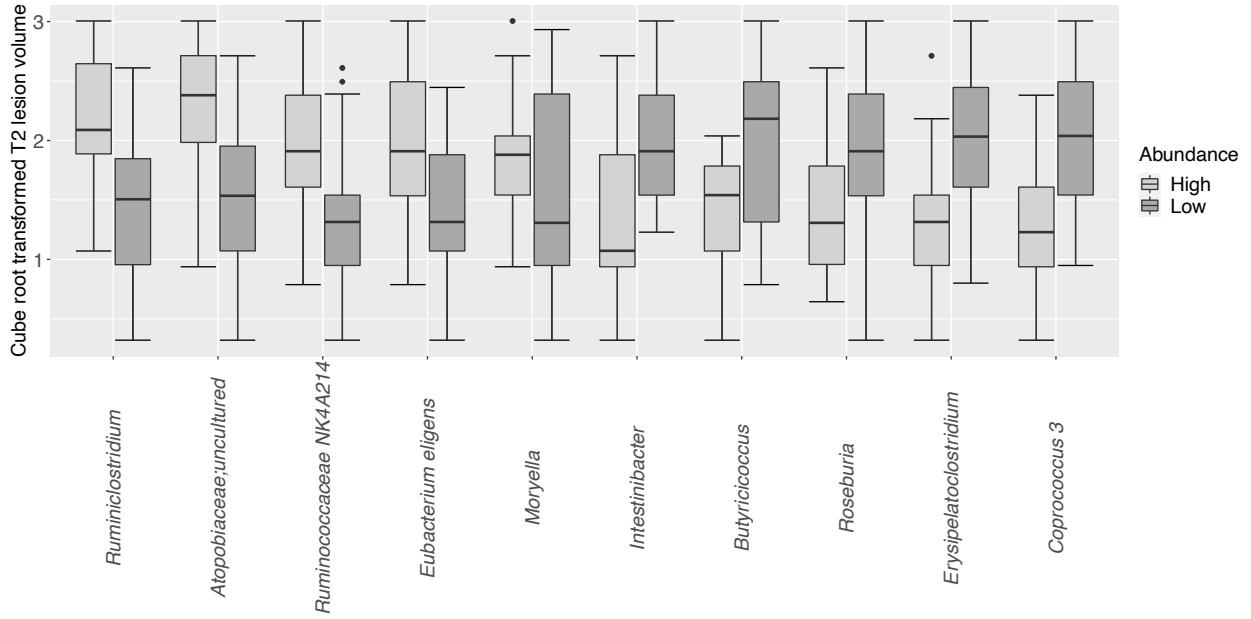
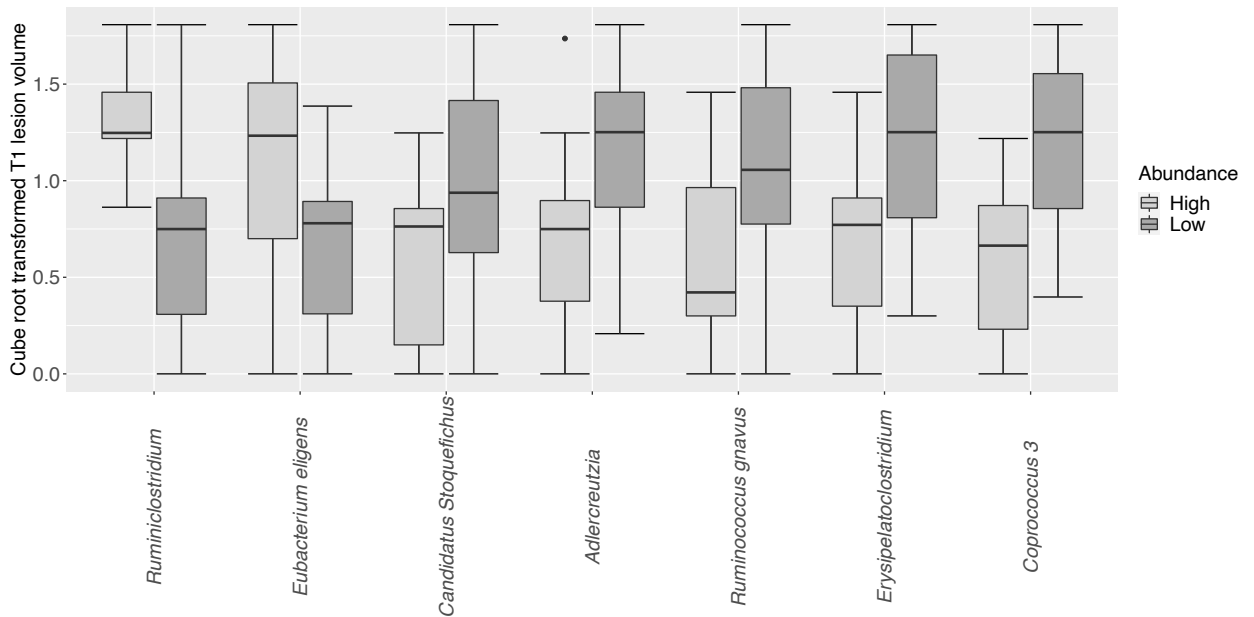


Figure S4 Boxplots of T2 and T1 lesion volumes by abundance level of the genera selected from the Lasso regression

a) T2 lesion volume



b) T1 lesion volume



Sensitivity Analysis

As a sensitivity analysis, we also considered the following variables in the Lasso regressions: age at MRI (continuous), sex, disease modifying drug exposure (ever vs never), country of residence (Canada vs USA), and Bristol Stool Scale (medium [types 3-5] vs hard or loose [types 1-2, 6-7]). The Bristol Stool Scale was grouped into two levels in the sensitivity analysis as only two participants' samples were categorized as loose (types 6-7).

Phylum-level gut microbiota

The selected taxa and corresponding β coefficients with 95% confidence intervals are reported in Table S5. Three of the five phyla associated with T2 lesion volume in the original analysis were also identified in this sensitivity analysis: *Firmicutes*, *Patescibacteria* and *Tenericutes*. The directions of the associations remain unchanged. Meanwhile, *Actinobacteria* and *Verrucomicrobia* were no longer selected.

All four phyla associated with T1 lesion volume in the original analysis were also selected in the sensitivity analysis with similar β coefficients. In addition, the phylum *Proteobacteria* was selected with a negative, albeit small, β coefficient (-0.03 [-0.33, 0.00]).

No additional covariates were selected in the analysis of T2 lesion volume; whereas, male sex was identified as being associated with a lower T1 lesion volume (β coefficients = -0.18 [-0.47, 0.00]).

Table S5 The β coefficients and 95% confidence intervals (CIs) of gut microbiota (phylum level) selected by the Lasso regression models with and without including additional covariates.

Taxon Name	T2 lesion Volume β coefficients (95% CIs)		T1 Lesion Volume β coefficients (95% CIs)	
	Without other covariates	With other covariates ^a	Without other covariates	With other covariates ^a
<i>Tenericutes</i>	0.37 (0.00, 0.85)	0.19 (0.00, 0.63)	0.25 (0.00, 0.60)	0.30 (0.00, 0.61)
<i>Verrucomicrobia</i>	0.07 (-0.03, 0.50)	-	-	-
<i>Actinobacteria</i>	-0.18 (-0.54, 0.00)	-	-0.22 (-0.55, 0.00)	-0.24 (-0.54, 0.00)
<i>Patesibacteria</i>	-0.21 (-0.63, 0.00)	-0.03 (-0.45, 0.00)	-0.18 (-0.49, 0.00)	-0.16 (-0.49, 0.00)
<i>Firmicutes</i>	-0.30 (-0.79, 0.00)	-0.25 (-0.63, 0.00)	-0.07 (-0.48, 0.00)	-0.04 (-0.47, 0.00)
<i>Proteobacteria</i>	-	-	-	-0.03 (-0.33, 0.00)

^a The other covariates include: age at MRI (continuous), sex, disease modifying drug exposure (ever vs never), country of residence (Canada vs USA), and Bristol Stool Scale (medium [types 3-5] vs hard or loose [types 1-2, 6-7]).

Genus-level gut microbiota

The selected taxa and corresponding β coefficients with 95% confidence intervals are reported in Table S6. The results remain unchanged for T2 lesion volume. For T1 lesion volume, two

genera were not selected: *Adlercreutzia* under the phylum *Actinobacteria* and *Candidatus Stoquefichus* under the phylum *Firmicutes*; both had close to null β coefficients in the original analysis. Again, the β coefficients of the selected genera were very similar to the original analysis.

Increase in age at MRI was associated with a slight increase in T2 lesion volume (β coefficients = 0.00 [0.00, 0.02] per year). No other covariates were selected in the analysis of T1 lesion volume.

Table S6 The β coefficients and 95% confidence intervals (CIs) of gut microbiota (genus level) selected by the Lasso regression models with and without including additional covariates.

Taxon Name	T2 lesion Volume β coefficients (95% CIs)		T1 Lesion Volume β coefficients (95% CIs)	
	Without covariates	With covariates ^a	Without covariates	With covariates ^a
<i>Ruminiclostridium</i>	0.10 (0.00, 0.42)	0.10 (0.00, 0.42)	0.18 (0.00, 0.34)	0.15 (0.00, 0.32)
<i>Atopobiaceae;</i> <i>Uncultured</i>	0.09 (0.00, 0.41)	0.09 (0.00, 0.41)	-	-
<i>Ruminococcaceae</i> <i>NK4A214</i>	0.06 (0.00, 0.31)	0.06 (0.00, 0.31)	-	-
<i>Eubacterium eligens</i>	0.03 (0.00, 0.23)	0.02 (0.00, 0.23)	0.04 (0.00, 0.20)	0.01 (0.00, 0.19)
<i>Moryella</i>	0.01 (0.00, 0.25)	0.00 (0.00, 0.25)	-	-
<i>Intestinibacter</i>	-0.05 (-0.35, 0.00)	-0.05 (-0.35, 0.00)	-	-
<i>Butyricicoccus</i>	-0.09 (-0.35, 0.00)	-0.09 (-0.35, 0.00)	-	-
<i>Roseburia</i>	-0.09 (-0.29, 0.00)	-0.09 (-0.28, 0.00)	-	-
<i>Erysipelatoclostridium</i>	-0.24 (-0.50, 0.00)	-0.24 (-0.50, 0.00)	-0.14 (-0.34, 0.00)	-0.12 (-0.33, 0.00)
<i>Coprococcus 3</i>	-0.31 (-0.54, 0.00)	-0.31 (-0.54, 0.00)	-0.28 (-0.52, 0.00)	-0.28 (-0.51, 0.00)
<i>Ruminococcus</i> <i>gnavus</i>	-	-	-0.09 (-0.30, 0.00)	-0.06 (-0.27, 0.00)
<i>Candidatus</i> <i>Stoquefichus</i>	-	-	0.00 (-0.22, 0.00)	-
<i>Adlercreutzia</i>	-	-	0.00 (-0.36, 0.00)	-

^a The other covariates include age at MRI (continuous), sex, disease modifying drug exposure (ever vs never), country of residence (Canada vs USA), and Bristol Stool Scale (medium [types 3-5] vs hard or loose [types 1-2, 6-7]).