Supplementary materials

Akkermansia muciniphila quantification by qPCR.

Because of the limitation of the V3fwd and V4rev primers to target the 16s rDNA of *A. municiphila*, we performed a quantitative PCR (qPCR) targeting specifically *A. muciniphila*. The qPCR was performed using the 16S rRNA primers (forward CAG-CAC-GTG-AAG-GTG-GGG-AC and reverse CCT-TGC-GGT-TGG-CTT-CAG-AT) for *A. muciniphila* detection and amplification. Total 16S rRNA was quantified using bacterial universal primers (forward CGG-TGA-ATA-CGT-TCC-CGG and reverse TAC-GGC-TAC-CTT-GTT-ACG-ACTT). qPCR was performed in 20µL reaction volumes containing respectively 250nM and 200nM primers, a 1x concentration of Takyon™ No Rox SYBR® MasterMix (Eurogentec), and 6µL of template. The procedures were done by pre-denaturation of DNA samples at 95°C for 10 min and amplification of the denatured DNA samples with 40 cycles at 95°C for 10 s and 60°C for 1 min. The fluorescent product was detected in the last step of each cycle, and the specific amplicon was verifying by melting curve performed after amplification. Data analysis was performed using Bio-Rad CFX Maestro software, Cq values were measured in duplicate and calculated using a regression method. For quantification of the total bacterial 16S rRNA, standard curves were prepared with 10-fold serial dilutions of plasmid and Cq values are inversely proportional to the 16s rRNA copy number. gDNA serving as a standard for the specific PCR of *A. muciniphila was calculated by the standard curve* of total bacteria. Data were expressed as $\log_{10}$ copies of 16S rRNA genes/ g of wet feces. The $\log_{10}$ transformed *A. muciniphila* was normalised to $\log_{10}$ total bacterial content and we refer to this measurement as *A. muciniphila* relative abundance. For 4 samples, no PCR amplification was detectable using the universal bacterial primer, and thus, the final number of fecal DNA samples for the analysis of *A. muciniphila* was 508.

PICRUSt2

Based on the 16S rRNA sequences, the pipeline of Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2) (1), version 2.4.1 implemented in the Galaxy software framework (2) was used for predicting functional abundance and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Briefly, PICRUST2 inserts 16S rRNA sequences into a reference tree from genomes in the Integrated Microbial Genomes (3) (IMG) database and predict gene families based on phylogeny. Based on the OTU table, PICRUST2 then predict abundances for each gene family corrected for predicted 16S rRNA copy number and KEGG pathway abundances were calculated through mappings of KEGG Ortholog (KO) groups to pathways.

Random Forests

The method is composed of two different steps. The first step eliminates all variables of low importance regarding the outcome according to the Random Forest score of importance. For the second step, we choose the interpretation approach to select variables most highly related to the response variable: we included all genera meeting a minimum threshold of 0.1% of relative abundance in at least 10% of samples. For this specific analysis, we applied the center log-ratio (CLR) transformation on absolute abundance in order to make the data more compatible with conventional statistical methods. (4) We also included the previously stated potential confounding factors: maternal prepregnancy BMI, maternal country of birth, gestational age, delivery mode, age, sex and human milk consumption. In order to assess the performance of random forests to select the variables most highly associated with the outcome, we used repeated cross-validation (3-fold, 100 repetitions) of random forests in the caret (5) R package (6). Briefly for this cross-validation, the outcome is predicted by repeatedly using a subset of samples as a training set and the remaining samples as the test set. As it is a 3-fold

cross validation, the dataset is randomly splitted into 3 smaller subsets and two subsets serve as training sets while the remaining one serves as test set. An iteration is performed when each subset has served as training set once and the procedure is repeated 100 times. We computed R^2 both with and without the confounding factors (Table S13). The same approach was used to assess the performance of the model for the selection of KEGG pathways most highly related of BMI z-score at 5 years (Table S13).

References

1. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, et al. PICRUSt2 for prediction of metagenome functions. Nat Biotechnol. 2020 Jun;38(6):685–8.

2. Escudié F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, et al. Frogs: find, rapidly, otus with galaxy solution. Berger B, editor. Bioinformatics. 2018 Apr 15;34(8):1287–94.

3. Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Grechkin Y, et al. IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res. 2012 Jan 1;40(D1):D115–22.

4. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional. Front Microbiol. 2017 Nov 15;8:2224.

5. Kuhn M. Building Predictive Models in *R* Using the caret Package. J Stat Softw [Internet]. 2008 [cited 2022 Dec 8];28(5). Available from: http://www.jstatsoft.org/v28/i05/

6. Genuer R, Poggi JM, Tuleau-Malot C. VSURF: An R Package for Variable Selection Using Random Forests. R J. 2015;7(2):19.

Supplementary figures and tables

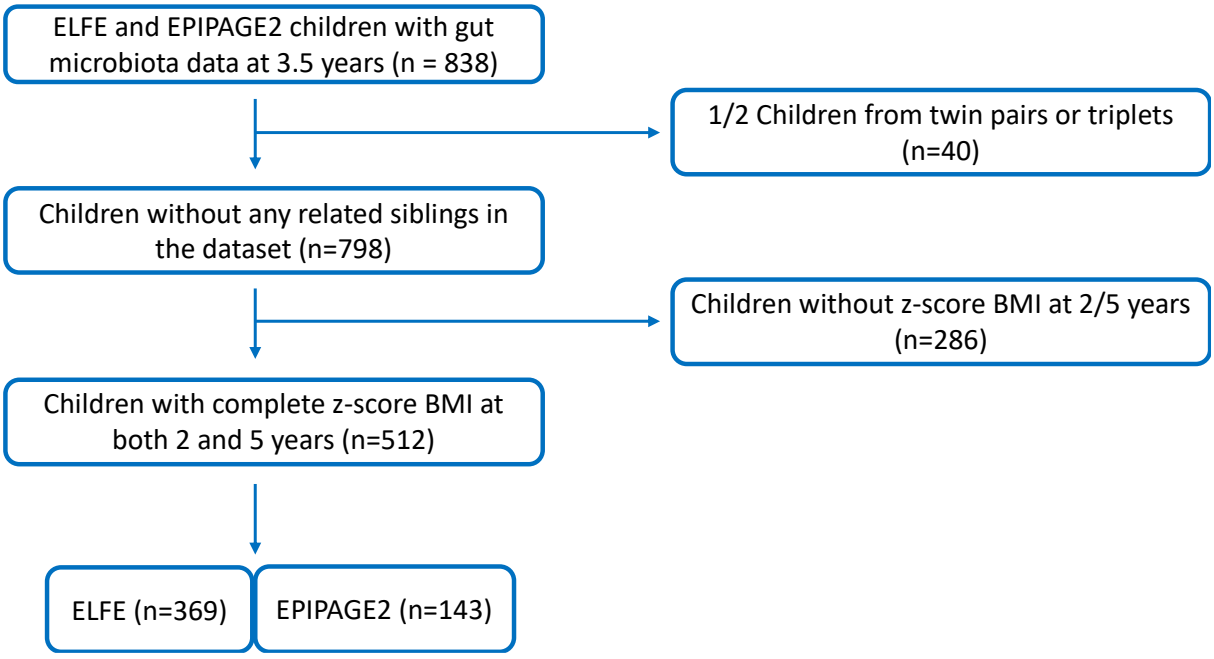

Figure S1. Flowchart of the study population.

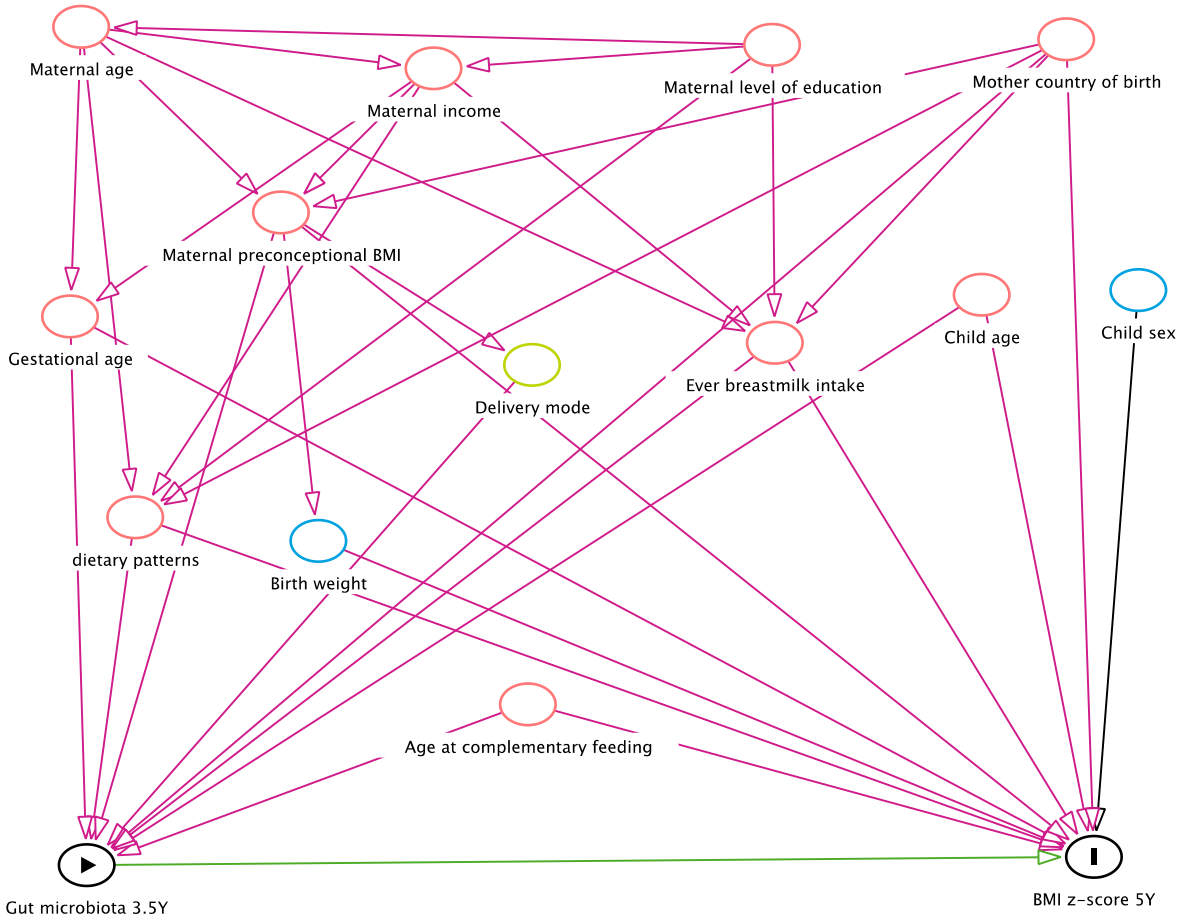

Figure S2. Directed Acyclic Graph created to design a multivariate model adjusted for a set of parsimonious potential confounding factors to assess the association between gut microbiota metric at 3.5 years and BMI z-score at 5 years.

Legend colors:

exposure

outcome

ancestor of exposure

ancestor of outcome

ancestor of exposure *and* outcome

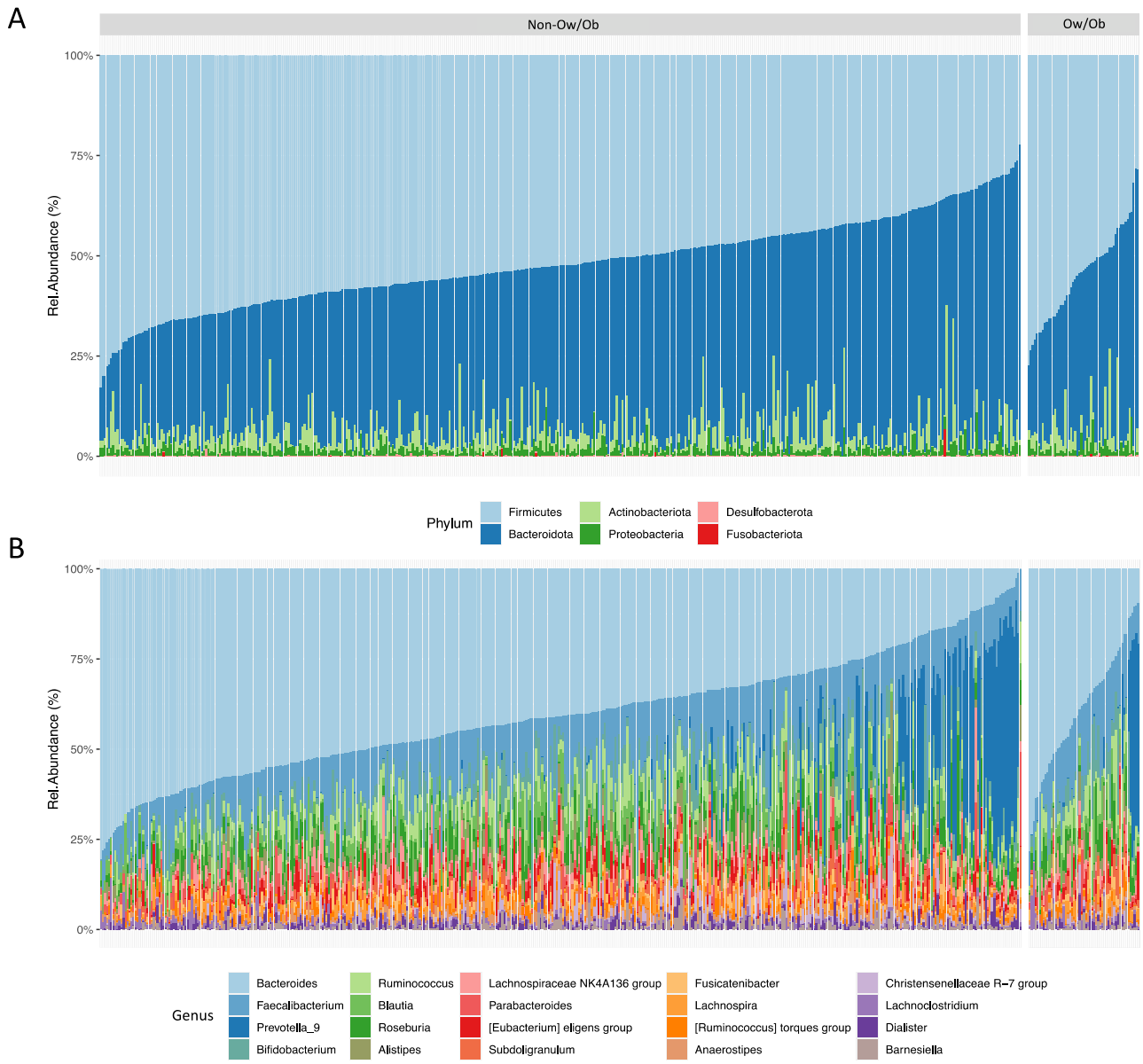

Figure S3 Children gut microbiota composition of the study population at the Phylum (A) and genus (B) levels according to their weight status.

Figure S4 shows the composition of the gut microbiota at both phylum and genus levels. The children gut microbiota was characterized by 6 phyla (*Firmicutes*, *Bacteroidetes*, *Actinobacteriota*, *Proteobacteria*, *Desulfobacterota*, and *Fusobacteriota*) where *Firmicutes* and *Bacteroidetes* were the most represented phyla. At the genus level, the most abundant genera were *Bacteroides*, *Faecalibacterium*, *Prevotella_9*, *Bifidobacterium* and *Ruminococcus*.

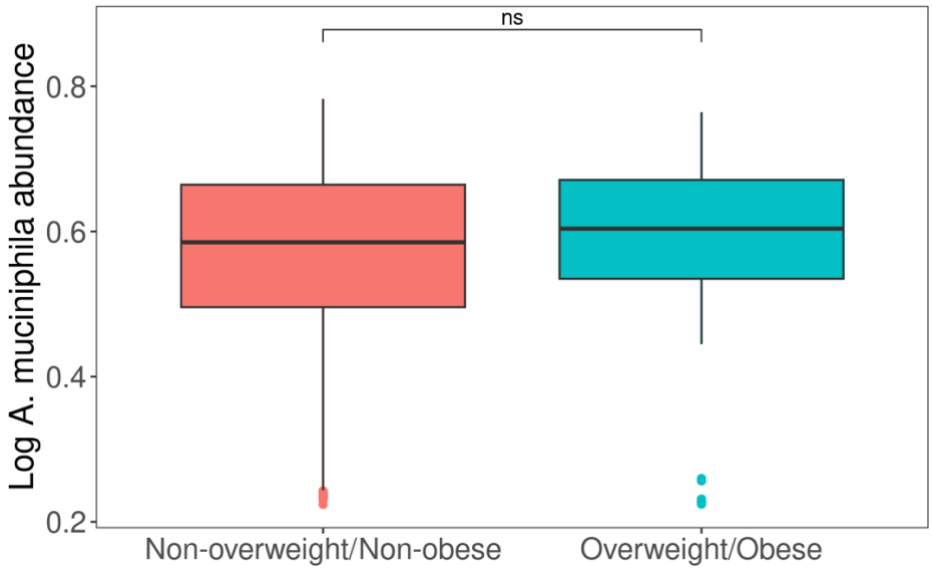

Figure S4. Relative abundance of *A. muciniphila* according to weight status at 5 years.
The boxplots show the smallest and largest values, 25 and 75% quartiles, the median, and outliers.
ns : Nonsignificant results according to Wilcoxon-Mann-Whitney test.

Table S1. PCA factor loading for dietary patterns at 2 years.

	PCA Factor loadings	
	Unhealthy	Healthy
Fruit juices	0.280	-0.245
Cereals	0.090	0.094
Cheese	**0.377**	0.129
Dairy products	0.271	0.197
Pasta	**0.432**	**0.455**
Cooked vegetables	**0.364**	**0.691**
Meat	**0.472**	0.245
Fresh fruits	0.283	0.241
Stewed fruits	0.204	**0.364**
Bread	**0.355**	0.202
Pastries	**0.402**	-0.283
Sweet beverages	0.285	**-0.430**
Fries	**0.400**	**-0.488**
Quiches	**0.370**	**-0.361**
Crudités	0.291	0.127
Charcuterie	**0.360**	**-0.469**
Fish	0.253	0.121
Eggs	0.252	-0.046
Candies	**0.333**	-0.170
Chips	0.272	**-0.312**

Factor loadings represent the correlations of all food items with a given dietary pattern.
In bold: Factor loadings >0.30 or < -0.30. We put a threshold of >0.30 for a positive and <-30 for a negative contribution to the dietary pattern.
PCA: Principal Component Analysis

Each participant was characterized with a score for each pattern: the higher the score, the greater their adherence to a given pattern. Patterns were defined according to their factor loadings as follows: (1) “unhealthy” where cheese, meat, pastries, chips, quiches were positively correlated; (2) “healthy” where cooked vegetables, pasta and stewed fruits were positively correlated while sweet beverages, fries, quiches and chips were negatively correlated (Table S1).

Table S2. Significance of each interaction term between gut microbiota metrics at 3.5 years found to be significantly associated with BMI z-score at 5 years and gestational age.

	Beta† (95% CI)	p-value
F/B ratio * Gestational age	-0.0005 (-0.2541, 0.2530)	>0.99
Chao1 estimate (Genus level) * Gestational age	0.003 (-0.010, 0.016)	0.70
Shannon index (Genus level) * Gestational age	-0.08 (-0.58, 0.42)	0.75
Genus		
Eggerthella * Gestational age	0.01 (-0.18, 0.20)	0.91
[Eubacterium] hallii group * Gestational age	0.01 (-0.12, 0.14)	0.92
Colidextribacter * Gestational age	0.0012 (-0.14, 0.14)	0.99
Ruminococcaceae CAG-352 * Gestational age	-0.01 (-0.11, 0.08)	0.82
Fusicatenibacter * Gestational age	0.02 (-0.07, 0.10)	0.71
[Eubacterium] ventriosum group * Gestational age	-0.07 (-0.21, 0.07)	0.32
KEGG pathways		
ko00140.Steroid.hormone.biosynthesis * Gestational age	-0.22, 0.15 (0.71)	0.71
ko00780.Biotin.metabolism * Gestational age	-0.18, 0.20 (0.92)	0.92
ko01040.Biosynthesis.of.unsaturated.fatty.acids * Gestational age	-0.28, 0.10 (0.37)	0.37
ko00531.Glycosaminoglycan.degradation * Gestational age	-0.16, 0.22 (0.74)	0.74
ko00520.Amino.sugar.and.nucleotide.sugar.metabolism * Gestational age	-0.10, 0.27 (0.38)	0.38

Beta coefficient and 95% confidence interval (CI) from multivariable linear regressions. We dichotomized the gestational age (GA) according to preterm status (<37 GA or ≥ 37 GA) and beta estimates are presented for "≥ 37 GA" compared to the reference group "< 37 GA". Each interaction term has been evaluated within its own model.
F/B: *Firmicutes/Bacteroidetes*
†Models are adjusted for maternal prepregnancy BMI, maternal country of birth, gestational age, delivery mode, age, sex and human milk consumption.

Table S3. Associations between the BMI z-score at 5 years and the relative abundance of *A. muciniphila* at 3.5 years quantified by qPCR.

N = 508	BMI z-score 5 years				Weight status (Ow/Ob vs non-Ow/Ob)			
	Beta[c] (95% CI)	p-value	Beta† (95% CI)	p-value	OR[c] (95% CI)	p-value	OR† (95% CI)	p-value
***A. muciniphila* relative abundance**	0.11 (-0.49, 0.71)	0.72	0.18 (-0.40, 0.77)	0.54	2.53 (0.32, 24.0)	0.39	3.86 (0.47, 38.1)	0.22

Beta coefficient and 95% confidence interval (CI) from multivariable linear regressions.
Odds ratio and 95% confidence interval (CI) from multivariable logistic regressions. The reference for the models is the non-Ow/Ob category defined by age and sex-specific BMI z-score WHO standard cut-offs.
†Models are adjusted for maternal prepregnancy BMI, maternal country of birth, gestational age, delivery mode, age, sex and human milk consumption.

Table S4. Associations between BMI z-score at 5 years and overall children gut microbiota composition at 3.5 years.

N=512		BMI z-score 5 years			
	Taxonomic level	Weighted Unifrac	Unweighted UniFrac	Bray Curtis	Omnibus
Unajusted	OTU	**0.014**	**0.015**	**0.014**	**0.043**
	Genus	**0.014**	**0.044**	**0.003**	**0.008**
Ajusted*	OTU	0.085	0.066	0.114	0.179
	Genus	0.087	0.129	**0.033**	0.093
Ajusted**	OTU	0.328	0.606	0.306	0.615
	Genus	0.356	0.458	0.299	0.589

Estimated p-values from unadjusted and adjusted Microbiome Regression-Based Kernel Association Tests (MiRKAT) of the weighted, unweighted UniFrac, and Bray-Curtis distances and BMI z-score at 5 years. Omnibus tests were also performed to simultaneously consider multiple distances.

*Models are adjusted for maternal prepregnancy BMI, maternal country of birth, gestational age, delivery mode, age, sex and human milk consumption.

** Models are adjusted for maternal prepregnancy BMI, maternal country of birth, gestational age, delivery mode, age, sex, human milk consumption, and BMI z-score at 2 years.

P-values in bold denote a statistical significance for p-values ≤ 0.05.

Table S5. Associations between BMI z-score at 5 years stratified by cohort and overall children gut microbiota composition at 3.5 years.

		BMI z-score 5 years							
		ELFE cohort (n = 369)				EPIFLORE cohort (n = 143)			
	Taxonomic level	Weighted Unifrac	Unweighted UniFrac	Bray Curtis	Omnibus	Weighted Unifrac	Unweighted UniFrac	Bray Curtis	Omnibus
Unajusted	OTU	**0.043**	0.064	0.096	0.103	0.758	0.118	0.292	0.273
	Genus	**0.040**	**0.049**	**0.013**	**0.041**	0.853	0.479	0.550	0.778
Ajusted*	OTU	**0.042**	0.087	0.100	0.100	0.909	0.309	0.578	0.587
	Genus	**0.041**	0.065	**0.027**	0.078	0.949	0.630	0.798	0.890
Ajusted†	OTU	**0.046**	0.077	0.105	0.107	"	"	"	"
	Genus	**0.044**	0.059	**0.027**	0.073	"	"	"	"
Ajusted**	OTU	0.202	0.443	0.396	0.406	0.970	0.550	0.631	0.844
	Genus	0.204	0.221	0.323	0.421	0.996	0.605	0.945	0.869

Estimated p-values stratified by cohort from unadjusted and adjusted Microbiome Regression-Based Kernel Association Tests (MiRKAT) of the weighted, unweighted UniFrac, and Bray-Curtis distances and BMI z-score at 5 years. Omnibus tests were also performed to simultaneously consider multiple distances.

*Models are adjusted for maternal prepregnancy BMI, maternal country of birth, gestational age, delivery mode, age, sex and human milk consumption.

†Models are adjusted for maternal prepregnancy BMI, maternal country of birth, gestational age, delivery mode, age, sex, human milk consumption, age at complementary feeding, and dietary patterns

** Models are adjusted for maternal prepregnancy BMI, maternal country of birth, gestational age, delivery mode, age, sex, human milk consumption, and BMI z-score at 2 years.

“ Models were not computed due to the absence of information for age at complementary feeding and dietary patterns for EPIFLORE cohort children.

P-values in bold denote a statistical significance for p-values ≤ 0.05.

Table S6. Associations between BMI z-score variation between 2 and 5 years and children gut alpha diversity measurements, and F/B ratio at 3.5 years.

N = 512	BMI z-score 5 years	
	Beta† (95% CI)	p-value
Chao1 estimate OTU level	-0.0004 (-0.0008, 0.0008)	0.92
Chao1 estimate Genus level	0.0003 (-0.004, 0.005)	0.91
Shannon index OTU level	0.03 (-0.14, 0.20)	0.71
Shannon index Genus level	0.02 (-0.15, 0.20)	0.78
Firmicutes/Bacteroidetes ratio	0.04 (-0.04, 0.12)	0.34

Beta coefficient and 95% confidence interval (CI) from multivariable linear regressions.
†Models are adjusted for maternal prepregnancy BMI, maternal country of birth, gestational age, delivery mode, age, sex, human milk consumption, and BMI z-score at 2 years.

Table S7. Associations between the BMI z-score at 2 years and the overall children gut microbiota composition at 3.5 years.

N = 512		BMI z-score 2 years			
	Taxonomic level	Weighted Unifrac	Unweighted UniFrac	Bray Curtis	Omnibus
Unajusted	**OTU**	**0.036**	**0.015**	**0.007**	**0.029**
	Genus	**0.033**	0.128	**0.009**	**0.046**
Ajusted*	**OTU**	0.167	**0.050**	**0.041**	0.108
	Genus	0.163	0.304	0.066	0.195

Estimated p-values from unadjusted and adjusted Microbiome Regression-Based Kernel Association Tests (MiRKAT) of the weighted, unweighted UniFrac, and Bray-Curtis distances and BMI z-score at 5 years. Omnibus tests were also performed to simultaneously consider multiple distances.
*Models are adjusted for maternal prepregnancy BMI, maternal country of birth, gestational age, delivery mode, age, sex and human milk consumption.
P-values in bold denote a statistical significance for p-values ≤ 0.05.

Table S8. Associations between the BMI z-score at 2 years and the children gut microbiota metrics at 3.5 years.

N = 512	BMI z-score 2 years	
	Beta† (95% CI)	p-value
Chao1 estimate OTU level	0.0003 (-0.001, 0.001)	0.60
Chao1 estimate Genus level	0.003 (-0.003, 0.01)	0.35
Shannon index OTU level	0.15 (-0.08, 0.38)	0.19
Shannon index Genus level	0.24 (0.01, 0.48)	**0.045**
Firmicutes/Bacteroidetes ratio	0.11 (0.002, 0.21)	**0.047**
Genus		
Eggerthella	-0.12 (-0.22, -0.02)	**0.014**
[Eubacterium] hallii group	0.05 (-0.01, 0.11)	0.12
Colidextribacter	-0.08 (-0.15, -0.01)	**0.027**
Lachnospiraceae UCG-001	-0.02 (-0.09, 0.05)	0.60
Bacteroides	-0.11 (-0.22, 0.00)	**0.041**
Ruminococcaceae CAG-352	0.00 (-0.05, 0.04)	0.87
Fusicatenibacter	0.04 (0.00, 0.08)	**0.041**
[Eubacterium] coprostanoligenes group	-0.01 (-0.06, 0.04)	0.74
Prevotella_9	0.01 (-0.01, 0.04)	0.32
[Eubacterium] ventriosum group	0.04 (-0.03, 0.11)	0.22
Anaerosporobacter	0.06 (-0.05, 0.16)	0.30
Sutterella	0.00 (-0.03, 0.04)	0.83
[Ruminococcus] gnavus group	-0.05 (-0.10, 0.01)	0.10
Lachnoclostridium	-0.09 (-0.19, 0.02)	0.10
Eisenbergiella	-0.03 (-0.09, 0.03)	0.28
KEGG pathways		
ko00281.Geraniol.degradation,	-0.05 (-0.14, 0.04)	0.30
ko00140.Steroid.hormone.biosynthesis	-0.12 (-0.21, -0.03)	**0.007**
ko03060.Protein.export	0.09 (0.00, 0.17)	0.054
ko01051.Biosynthesis.of.ansamycins	0.04 (-0.05, 0.13)	0.40
ko01040.Biosynthesis.of.unsaturated.fatty.acids	0.07 (-0.02, 0.16)	0.11
ko00780.Biotin.metabolism	-0.10 (-0.19, -0.01)	**0.025**
ko01053.Biosynthesis.of.siderophore.group.nonribosomal.peptides	-0.03 (-0.12, 0.06)	0.51
ko00720.Carbon.fixation.pathways.in.prokaryotes	-0.01 (-0.10, 0.08)	0.87
ko04210.Apoptosis	-0.05 (-0.14, 0.04)	0.25
ko00900.Terpenoid.backbone.biosynthesis	0.12 (0.03, 0.20)	**0.008**
ko00260.Glycine..serine.and.threonine.metabolism	0.05 (-0.04, 0.14)	0.26
ko00680.Methane.metabolism	-0.03 (-0.12, 0.06)	0.50
ko00400.Phenylalanine..tyrosine.and.tryptophan.biosynthesis	0.05 (-0.04, 0.14)	0.24
ko00521.Streptomycin.biosynthesis	-0.05 (-0.13, 0.04)	0.30
ko00531.Glycosaminoglycan.degradation	-0.11 (-0.19, -0.02)	**0.018**
ko00520.Amino.sugar.and.nucleotide.sugar.metabolism	-0.11 (-0.20, -0.02)	**0.015**
ko03450.Non.homologous.end.joining	-0.05 (-0.14, 0.04)	0.29

Beta coefficient and 95% confidence interval (CI) from multivariable linear regressions.
†Models are adjusted for maternal prepregnancy BMI, maternal country of birth, gestational age, delivery mode, age, sex and human milk consumption.
P-values in bold denote a statistical significance for p-values ≤ 0.05.

Table S9. Associations between the BMI z-score at 3.5 years and the overall ELFE children gut microbiota composition at 3.5 years.

N = 212		BMI z-score 3.5 years			
	Taxonomic level	**Weighted Unifrac**	**Unweighted UniFrac**	**Bray Curtis**	**Omnibus**
Unajusted	**OTU**	0.102	0.500	0.608	0.202
	Genus	0.099	0.431	0.068	0.141
Ajusted*	**OTU**	0.084	0.533	0.500	0.174
	Genus	0.081	0.474	**0.044**	0.099
Ajusted†	**OTU**	0.093	0.524	0.514	0.191
	Genus	0.089	0.446	**0.047**	0.107

Estimated p-values from unadjusted and adjusted Microbiome Regression-Based Kernel Association Tests (MiRKAT) of the weighted, unweighted UniFrac, and Bray-Curtis distances and BMI z-score at 5 years. Omnibus tests were also performed to simultaneously consider multiple distances.

†Models are adjusted for maternal prepregnancy BMI, maternal country of birth, gestational age, delivery mode, age, sex and human milk consumption.

* Models are adjusted for maternal prepregnancy BMI, maternal country of birth, gestational age, delivery mode, age, sex, human milk consumption, age at complementary feeding and dietary pattern scores

Table S10. Associations between BMI z-score at 3.5 years and ELFE children gut microbiota metrics at 3.5 years.

N = 212	BMI z-score 3.5 years			
	Beta† (95% CI)	**p-value**	**Beta* (95% CI)**	**p-value**
Chao1 estimate OTU level	0.0002 (-0.001, 0.002)	0.80	0.0002 (-0.001, 0.002)	0.83
Chao1 estimate Genus level	-0.0021 (-0.01, 0.01)	0.70	-0.002 (-0.01, 0.02)	0.68
Shannon index OTU level	0.06 (-0.30, 0.43)	0.74	0.06 (-0.31, 0.43)	0.76
Shannon index Genus level	0.17 (-0.20, 0.53)	0.37	0.16 (-0.21, 0.53)	0.39
Firmicutes/Bacteroidetes ratio	0.17 (0.001, 0.34)	**0.048**	0.17 (0.003, 0.34)	0.054
Genus				
Eggerthella	-0.19 (-0.35, -0.03)	**0.018**	-0.19 (-0.35, -0.03)	**0.019**
[Eubacterium] hallii group	0.14 (0.05, 0.23)	**0.003**	0.14 (0.05, 0.23)	**0.003**
Colidextribacter	-0.03 (-0.15, 0.09)	0.61	-0.03 (-0.15, 0.09)	0.61
Lachnospiraceae UCG-001	0.02 (-0.09, 0.14)	0.70	0.02 (-0.09, 0.14)	0.70
Bacteroides	-0.19 (-0.39, 0.01)	**0.061**	-0.19 (-0.39, 0.01)	**0.063**
Ruminococcaceae CAG-352	0.01 (-0.06, 0.09)	0.72	0.01 (-0.06, 0.09)	0.74
Fusicatenibacter	0.08 (0.01, 0.14)	**0.026**	0.08 (0.01, 0.14)	**0.028**
[Eubacterium] coprostanoligenes group	0.04 (-0.04, 0.12)	0.38	0.04 (-0.04, 0.12)	0.36
Prevotella_9	0.02 (-0.02, 0.07)	0.35	0.02 (-0.02, 0.07)	0.32
[Eubacterium] ventriosum group	0.10 (-0.02, 0.21)	0.094	0.10 (-0.02, 0.21)	0.10
Anaerosporobacter	0.07 (-0.09, 0.23)	0.39	0.07 (-0.09, 0.23)	0.41
Sutterella	-0.02 (-0.08, 0.05)	0.62	-0.01 (-0.08, 0.05)	0.64
[Ruminococcus] gnavus group	-0.07 (-0.16, 0.02)	0.13	-0.07 (-0.16, 0.02)	0.12
Lachnoclostridium	-0.19 (-0.37, 0.00)	0.053	-0.02 (-0.12, 0.08)	0.68
Eisenbergiella	-0.02 (-0.12, 0.08)	0.66	-0.02 (-0.12, 0.08)	0.68
KEGG pathways				
ko00281.Geraniol.degradation,	-0.16 (-0.32, 0.00)	0.054	-0.16 (-0.33, 0.00)	**0.049**
ko00140.Steroid.hormone.biosynthesis	-0.21 (-0.35, -0.06)	**0.006**	-0.21 (-0.36, -0.06)	**0.006**
ko03060.Protein.export	0.11 (-0.04, 0.26)	0.15	0.11 (-0.04, 0.26)	0.15
ko01051.Biosynthesis.of.ansamycins	0.09 (-0.06, 0.23)	0.24	0.09 (-0.06, 0.24)	0.22
ko01040.Biosynthesis.of.unsaturated.fatty.acids	0.18 (0.03, 0.33)	**0.018**	0.18 (0.03, 0.33)	**0.019**
ko00780.Biotin.metabolism	-0.11 (-0.24, 0.02)	0.11	-0.11 (-0.25, 0.02)	0.11
ko01053.Biosynthesis.of.siderophore.group.nonribosomal.peptides	-0.15 (-0.32, 0.03)	0.10	-0.15 (-0.33, 0.03)	0.10
ko00720.Carbon.fixation.pathways.in.prokaryotes	-0.13 (-0.29, 0.02)	0.093	-0.14 (-0.30, 0.02)	**0.085**
ko04210.Apoptosis	-0.10 (-0.24, 0.05)	0.19	-0.10 (-0.25, 0.05)	0.19
ko00900.Terpenoid.backbone.biosynthesis	0.10 (-0.04, 0.24)	0.15	0.10 (-0.04, 0.24)	0.15
ko00260.Glycine..serine.and.threonine.metabolism	0.04 (-0.10, 0.18)	0.55	0.04 (-0.10, 0.18)	0.55
ko00680.Methane.metabolism	0.03 (-0.10, 0.17)	0.63	0.04 (-0.10, 0.18)	0.62
ko00400.Phenylalanine..tyrosine.and.tryptophan.biosynthesis	0.12 (-0.02, 0.26)	0.093	0.12 (-0.02, 0.26)	0.086
ko00521.Streptomycin.biosynthesis	-0.05 (-0.20, 0.09)	0.48	-0.06 (-0.20, 0.09)	0.46
ko00531.Glycosaminoglycan.degradation	-0.17 (-0.31, -0.03)	**0.020**	-0.17 (-0.32, -0.03)	**0.021**
ko00520.Amino.sugar.and.nucleotide.sugar.metabolism	-0.07 (-0.21, 0.06)	0.30	-0.07 (-0.21, 0.06)	0.30
ko03450.Non.homologous.end.joining	0.02 (-0.16, 0.19)	0.85	0.02 (-0.16, 0.19)	0.85

Beta coefficient and 95% confidence interval (CI) from multivariable linear regressions.
†Models are adjusted for maternal prepregnancy BMI, maternal country of birth, gestational age, delivery mode, age, sex and human milk consumption.
*Models are adjusted for maternal prepregnancy BMI, maternal country of birth, gestational age, delivery mode, age, sex, human milk consumption, age at complementary feeding and dietary pattern scores
P-values in bold denote a statistical significance for p-values ≤ 0.05.

Table S11. Complete cases sensitivity analysis of associations between BMI z-score at 5 years and overall children gut microbiota composition at 3.5 years.

N=440		Zscore BMI 5 years			
	Taxonomic level	**Weighted Unifrac**	**Unweighted UniFrac**	**Bray Curtis**	**Omnibus**
Unajusted	**OTU**	**0.018**	**0.021**	**0.045**	**0.041**
	Genus	**0.017**	**0.038**	**0.002**	**0.011**
Ajusted*	**OTU**	0.072	0.081	0.273	0.166
	Genus	0.066	0.121	**0.022**	0.057

Estimated p-values from unadjusted and adjusted Microbiome Regression-Based Kernel Association Tests (MiRKAT) of the weighted, unweighted UniFrac, and Bray-Curtis distances and BMI z-score at 5 years. Omnibus tests were also performed to simultaneously consider multiple distances.

*Models are adjusted for maternal prepregnancy BMI, maternal country of birth, gestational age, delivery mode, age, sex and human milk consumption.

P-values in bold denote a statistical significance for p-values ≤ 0.05.

Table S12. Complete cases sensitivity analysis of associations between BMI z-score and weight status at 5 years and children gut microbiota metrics at 3.5 years.

N=440	BMI z-score 5 years		Weight status (Ow/Ob vs non-Ow/Ob)	
	Beta† (95% CI)	p-value	OR† (95% CI)	p-value
Chao1 estimate OTU level	0.0002 (-0.0009, 0.001)	0.75	1.00 (0.99, 1.01)	0.58
Chao1 estimate Genus level	0.0024 (-0.004, 0.009)	0.47	1.00 (0.97, 1.02)	0.73
Shannon index OTU level	0.14 (-0.10, 0.37)	0.25	0.71 (0.32, 1.58)	0.40
Shannon index Genus level	0.20 (-0.04, 0.44)	0.10	0.87 (0.38, 1.98)	0.73
Firmicutes/Bacteroidetes ratio	0.13 (0.02, 0.25)	**0.021**	1.42 (0.98, 2.01)	0.066
Genus				
Eggerthella	-0.14 (-0.24, -0.04)	**0.007**	0.73 (0.44, 1.09)	0.14
[Eubacterium] hallii group	0.1 (0.04, 0.16)	**<0.001**	1.17 (0.96, 1.46)	0.13
Colidextribacter	-0.07 (-0.15, 0.00)	**0.046**	0.97 (0.75, 1.26)	0.83
Lachnospiraceae UCG-001	-0.05 (-0.13, 0.02)	0.17	0.98 (0.74, 1.27)	0.89
Bacteroides	-0.12 (-0.23, -0.01)	**0.033**	0.93 (0.65, 1.39)	0.72
Ruminococcaceae CAG-352	-0.07 (-0.12, -0.02)	**0.003**	0.71 (0.49, 0.91)	**0.004**
Fusicatenibacter	0.05 (0.01, 0.09)	**0.015**	1.11 (0.95, 1.32)	0.18
[Eubacterium] coprostanoligenes group	0.00 (-0.04, 0.05)	0.86	1 (0.85, 1.19)	0.99
Prevotella_9	0.02 (-0.01, 0.05)	0.11	1.02 (0.92, 1.13)	0.64
[Eubacterium] ventriosum group	0.06 (-0.01, 0.13)	**0.07**	1.11 (0.87, 1.43)	0.41
Anaerosporobacter	0.05 (-0.05, 0.16)	0.30	1.2 (0.85, 1.64)	0.29
Sutterella	0.01 (-0.03, 0.05)	0.61	0.98 (0.86, 1.12)	0.73
[Ruminococcus] gnavus group	-0.05 (-0.11, 0.002)	0.061	1.1 (0.91, 1.32)	0.33
Lachnoclostridium	-0.09 (-0.20, 0.01)	0.091	1.01 (0.70, 1.51)	0.95
Eisenbergiella	-0.04 (-0.10, 0.02)	0.24	1.04 (0.84, 1.28)	0.70
KEGG pathways				
ko00281.Geraniol.degradation,	-0.09 (-0.18, 0.00)	**0.047**	1.15 (0.84, 1.56)	0.36
ko00140.Steroid.hormone.biosynthesis	-0.12 (-0.21, -0.03)	**0.011**	0.96 (0.69, 1.32)	0.82
ko03060.Protein.export	0.06 (-0.03, 0.14)	0.23	0.98 (0.71, 1.36)	0.92
ko01051.Biosynthesis.of.ansamycins	0.06 (-0.02, 0.15)	0.16	1.08 (0.79, 1.48)	0.64
ko01040.Biosynthesis.of.unsaturated.fatty.acids	0.10 (0.01, 0.19)	**0.024**	0.98 (0.71, 1.34)	0.88
ko00780.Biotin.metabolism	-0.11 (-0.19, -0.02)	**0.016**	0.89 (0.64, 1.22)	0.46
ko01053.Biosynthesis.of.siderophore.group.nonribosomal.peptides	-0.09 (-0.17, 0.00)	0.054	0.94 (0.65, 1.28)	0.72
ko00720.Carbon.fixation.pathways.in.prokaryotes	-0.07 (-0.16, 0.02)	0.12	0.92 (0.67, 1.27)	0.63
ko04210.Apoptosis	-0.05 (-0.13, 0.04)	0.30	0.91 (0.66, 1.24)	0.56
ko00900.Terpenoid.backbone.biosynthesis	0.10 (0.01, 0.19)	**0.029**	0.94 (0.68, 1.29)	0.72
ko00260.Glycine..serine.and.threonine.metabolism	0.02 (-0.09, 0.09)	0.93	0.89 (0.64, 1.22)	0.47
ko00680.Methane.metabolism	-0.02 (-0.11, 0.07)	0.71	0.79 (0.56, 1.09)	0.15
ko00400.Phenylalanine..tyrosine.and.tryptophan.biosynthesis	0.08 (-0.01, 0.17)	0.071	0.92 (0.67, 1.27)	0.62
ko00521.Streptomycin.biosynthesis	-0.04 (-0.13, 0.04)	0.34	1.02 (0.73, 1.39)	0.92
ko00531.Glycosaminoglycan.degradation	-0.12 (-0.20, -0.03)	**0.010**	1.02 (0.73, 1.40)	0.91
ko00520.Amino.sugar.and.nucleotide.sugar.metabolism	-0.12 (-0.21, -0.03)	**0.007**	1.05 (0.76, 1.44)	0.76
ko03450.Non.homologous.end.joining	-0.03 (-0.12, 0.05)	0.45	1.00 (0.72, 1.34)	0.99

Beta coefficient and 95% confidence interval (CI) from multivariable linear regressions.
Odds ratio and 95% confidence interval (CI) from multivariable logistic regressions. The reference for the models is the non-Ow/Ob category defined by age and sex-specific BMI z-score WHO standard cut-offs.
†Models are adjusted for maternal prepregnancy BMI, maternal country of birth, gestational age, delivery mode, age, sex and human milk consumption.
P-values in bold denote a statistical significance for p-values ≤ 0.05.

Table S13. Amount of variation in BMI z-score at 5 years explained by the selected features of the VSURF procedure.

Model		R^2 (%)	RMSE
Genus			
Random Forest of all genera	Unadjusted	2.34 (0.56-4.12)	0.94 (0.90-0.97)
	Ajusted*	2.75 (0.74-4.77)	0.93 (0.89-0.97)
Random Forest of VSURF selected genera	Unadjusted	10.92 (6.06-12.89))	0.89 (0.85-0.95)
	Ajusted*	38.72 (34.18-43.26)	0.74 (0.69-0.78)
KEGG pathways			
Random Forest of KEGG pathways	Unadjusted	0.81 (-0.09-1.73)	0.94 (0.91-0.99)
	Ajusted*	1.12 (-0.08-2.32)	0.94 (0.91-0.98)
Random Forest of VSURF selected KEGG pathways	Unadjusted	3.12 (0.66-5.58)	0.93 (0.88-0.98)
	Ajusted*	3.99 (1.07-6.89)	0.92 (0.87-0.98)

This table shows the model performance metrics (R^2 (95% confidence interval), or the percent of the variation explained in the outcome and the root mean squared error [RMSE]) of random forests to predict the BMI z-score at 5 years based on abundance of gut microbiota genera and KEGG pathways. These results are from 3-fold cross validation of 100 repetitions.