SUPPLEMENTARY MATERIALS

1.1 *Per Protocol Analysis*

We found that one participant had fallen below 50% adherence to treatment allocation as specified in the per protocol analysis. Therefore, we excluded this participant from the ITT analysis and repeated the analysis for the primary outcome. These data showed that using the difference in MRS-measured liver fat (MRS baseline - MRS end of study) as the outcome, and including treatment allocation and MRS-measured liver fat at baseline as independent factors in the multivariable regression model, the β coefficient for treatment allocation was -2.5; 95%CI -7.7, 2.5; p=0.30. When the results of the aforementioned regression model were further adjusted for age, sex, weight difference and weight at baseline, the β coefficient for treatment allocation was -4.4; 95% CI -0.07, 0.3; p=0.03. N.B.: the negative unstandardized β coefficient we found in this regression model indicates that there is a reduction in MRS-measured liver fat in the placebo group versus the synbiotic group.

1.2 *Biochemical measurements, body composition, energy expenditure and measurements of liver fibrosis*

All measurements were undertaken at baseline and at the end of the study. Fibrosis markers, including hyaluronic acid (HA), procollagen-III N-terminal propeptide (PIIINP) and tissue inhibitor of matrix metalloproteinase (TIMP-1), were measured. We generated two different histologically-validated liver fibrosis scores (27-29). The first of these scores comprised measurement of serum HA, PIIINP and TIMP-1 concentrations (i.e., the Enhanced Liver Fibrosis [ELF] score) (28) and the second validated algorithmically-derived score (i.e., the NAFLD fibrosis score) used age (years), body mass index (BMI), impaired fasting glucose/ diabetes (yes/no),

alanine aminotransferase-to-aspartate aminotransferase (ALT/AST) ratio, platelet count and serum albumin concentration (29). Body composition was assessed by Dual Energy X Ray Absorptiometry (DEXA or DXA), and liver stiffness measurement by Vibration-Controlled Transient Elastograpy (VCTE^{TM}) was measured by Fibroscanner (Echosens) (27). Body composition was measured using simple anthropometric measurements and physical activity was assessed using armband monitoring. Any change in diet during the trial was assessed by food frequency questionnaire (FFQ) (27).

1.3 *Liver fat measurement*

Measurement of liver fat content was undertaken at baseline and at the end of the study using proton magnetic resonance spectroscopy (MRS). Three 20 × 20 × 20 mm^3 spectroscopic voxels were positioned within segments 3 (inferior sub-segment of the lateral segment), 5 (inferior sub-segment of the anterior segment) and 8 (superior sub-segment of the anterior segment) of the liver, avoiding major blood vessels, intra-hepatic bile ducts and the lateral margin of the liver. For the second visit scan, these voxel positions were copied from the first scan for each participant, to ensure measurements at the end of the study were taken in the same regions of the liver, as at baseline. Values for the lipid and water peak integrals were produced for each voxel and recorded for each participant (27). For each participant to test the effect of the intervention on liver fat percentage we calculated the mean of the three 20 × 20 × 20 mm^3 spectroscopic voxels for liver fat percentage at baseline and end of study.

1.4 *Faecal microbiota analysis*

Faecal genomic DNA was extracted using the QIAamp DNA Stool Mini Kit, including a bead-beating step. Total bacteria, *Bifidobacterium* spp. and *Bifidobacterium animalis* levels were

quantified using qPCR as previously described (30). The samples were PCR-enriched for the V5-V6 region of the 16S rRNA gene and then underwent a library tailing PCR as described in Pötgens et al. (31). The amplicons were purified, quantified and sequenced using an Illumina MiSeq to produce 2 x 300 bp sequencing products. Initial quality-filtering of the reads was conducted with the Illumina Software, yielding an average of 169,858 pass-filter reads per sample. Quality scores were visualized and reads were trimmed to 220 bp (R1) and 200 bp (R2). The reads were merged with the merge-Illumina-pairs application (32). For all samples but two, a subset of 26,000 reads was randomly selected using Mothur 1.32.1 (33). The UPARSE pipeline implemented in USEARCH v10.0.240 (34) was used to further process the sequences. Putative chimaeras were removed using the *cluster_otus* command. Non-chimeric sequences were subjected to taxonomic classification using the RDP MultiClassifier 1.1 from the Ribosomal Database Project (35). The phylotypes were computed as percent proportions based on the total number of sequences in each sample. Raw sequences can be found in the SRA database (project ID: PRJNA559052).

1.5

1.6

1.7 ***Inclusion and exclusion criteria***

Inclusion and exclusion criteria have been described previously (33). Briefly, the inclusion criteria for participation in the study were: 1) diagnosis of liver fat on normal clinical grounds with either histological confirmation of NAFLD or imaging evidence of liver fat with exclusion of other liver conditions causing liver fat accumulation, and 2) alcohol consumption ≤14 units/week for women ≤ 21 units/week for men (34). Exclusion criteria included: 1) abdominal surgery, 2) three or more courses of broad-spectrum antibiotics in the year before enrolment that may change gut

microbiota (35), and 3) consumption of probiotic foods or supplements, within the two months preceding enrolment.

1.8

1.9 *Intestinal permeability assessment*

The lactulose-mannitol test was used to measure the intestinal permeability before and after the intervention. After an overnight fast, participants were asked to drink a solution containing 10 g of lactulose and 5 g of mannitol in 35 ml of water (1300 mOsml/L). Urine was collected (over the 6 hours after the drink) into plastic containers with 1 ml of chlorexidine 2% as preservative, to prevent bacterial degradation of sugars and a 10 ml urine sample was stored at -20°C until assayed. Lactulose and mannitol concentrations in the urine samples were then measured on an auto-analyser (Konelab 20; Thermo Scientific, Swedesboro, NJ, USA), using enzymatic colorimetric techniques as described previously (37, 38). All reagents, unless stated otherwise, were purchased from Sigma Aldrich (Poole, UK) and BioCatalysts (Nantgarw, Wales). The results of the lactulose-mannitol test are expressed as a ratio of percentage of urinary excretion of lactulose to the percentage of urinary excretion of mannitol in relation to the amount of each probe consumed (L/M ratio). For the lactulose test, 10 lactulose assays on a urine sample containing a lactulose concentration of 0.58 mmol/L were carried out and the results of the assessment showed an intra-assay precision of 1.9%; while for the mannitol test, four mannitol assays on a urine sample containing a mannitol concentration of 1.5 mmol/L, the intra-assay precision was 0.5%.

1.10

1.11 *Faecal short-chain fatty acid analysis*

The preparation and analysis of short-chain fatty acids in faecal samples was carried out using methods developed in a previous study (39) with slight modifications. Freshly voided or frozen faecal samples were stored at -80ºC prior to analysis. Faecal slurry supernatants were used to determine the faecal concentrations of acetic acid, propionic acid and butyric acid by gas chromatography. Faecal slurry was prepared by dilution in PBS (0.1 M; pH 7.0) and vortexed with 2 g of 3 mm diameter glass beads (VWR) and then doubly centrifuged to remove particulate matter (1500 g, 2 min; 13000 g 10min). The resulting supernatant was collected and stored at -80ºC prior to further processing. To a 450 ul aliquot of faecal slurry supernatant, 50 ul of an internal standard (1 mM of 2-ethyl butyric acid in 1 M formic acid) and 1 ml of diethyl ether were added and the sample centrifuged at 3000 g for 10 min. 400 ul of the resulting upper ether layer was combined with 50 ul of N-tert-butyldimethysilyl-N-methytrifluoroacetamide, heated at 80ºC for 20 min in a water bath and then left at room temperature for a minimum of 48 h to allow derivatisation. Samples were analysed on an Agilent 6890 Series GC system (HP, Crawley, West Sussex, UK) fitted with a FameWax polyethylene glycol capillary column (30 m x 0.25 µm x 0.25 mm) (Thames Restek) and flame ionisation detector. The carrier gas, helium, was delivered at a flow rate of 2 ml/min. The head pressure was set at 14.5psi with split injection (split ratio of 2:1 used). Run conditions were: initial temperature 40°C, ramp 10°C/min to 150°C, hold for 3 min, ramp 20ºC/min to 225°C, hold for 10 min. Peaks were integrated using Agilent ChemStation software (Agilent Technologies, Oxford, UK). Short-chain fatty acids were identified by comparison with standards, and quantified relative to the internal standard. Data are presented as short-chain fatty acid concentrations in umol/g wet weight faeces.

1.12 *Serum lipopolysaccharide measurement*

Chromogenic limulus amoebocyte lysate (LAL) assays were carried out using Pierce™ Chromogenic Endotoxin Quant Kit (Thermo Fisher Scientific). All samples were processed within the same kit lot number. Single-use, pyrogen-free pipette tips, 96-well microplates and microcentrifuge tubes were used, and experiments were carried out in a class II biosafety cabinet (Thermo Electron Corporation). Serum samples were thawed and reagents were equilibrated at room temperature for 1 h before use, and the 96-well microplate was maintained at 37±1°C throughout the assay procedure using a digital dry bath (Labnet D1200 Accublock). *Escherichia coli* endotoxin standard was reconstituted by adding 2 mL of sterile, endotoxin-free Dulbecco's PBS (Sigma Aldrich) to a final concentration of 10 EU/mL. The solution was placed on a plate shaker for 15 minutes and then serially-diluted with sterile, endotoxin-free Dulbecco's PBS to prepare 0.01-0.1 EU/mL standards. Serum samples were diluted 1:10 with sterile, endotoxin-free Dulbecco's PBS and assayed in duplicate. The chromogenic assay was carried out in accordance with manufacturer instructions, and the optical density (OD) was measured at 405 nm on a microplate reader (Thermo Labsystems Multiskans EX). The average absorbance of the blank replicates was subtracted from the average absorbance of all individual standards and sample replicates. Concentrations of endotoxin in the samples were calculated using a second order polynomial (quadratic) curve-fit on GraphPad Prism v.8, and results are expressed in EU/mL serum.