# SUPPLEMENTAL MATERIAL FOR

**A wholegrain-rich diet reduces body weight and systemic low-grade inflammation without inducing major changes of the gut microbiome: A randomised cross-over trial**

**SUPPLEMENTARY METHODS**

# SUPPLEMENTARY METHODS

## Analyses of study product composition

*Resistant starch, arabinoxylan and alkylresorcinol content in food products provided to the participants*

The details of the analysis of resistant starch content in the study products are described elsewhere.[1] The determination was performed using a downscaled modified procedure according to Åkerberg *et al.*[2], including an *in vivo* chewing step followed by an *in vitro* digestion procedure. Arabinoxylan content was determined by quantifying the reducing sugars arabinose and xylose by gas-liquid chromatography as described by Courtin *et al.*[3] The concentrations of alkylresorcinol homologs (C15:0, C17:1, C17:0, C19:1, C19:0, C21:0, C21:1, C23:0, C25:1, and C25:0) in the products were analysed using ultra high-pressure liquid chromatography with fluorescence detection as described elsewhere.[4]

*Average monosaccharide composition of the two intervention diets*

An average meal representing each intervention diet was prepared based on average daily intake among study participants reported in the study diaries. The samples were freeze dried and milled to a ≤1 mm particle size. In order to determine the amount of starch glucose and non-starch glucose, subsamples of the two diets where de-starched prior to hydrolysis. Samples were treated with 0.2% v/w Termamyl® SC in 85 min at 70 °C in 50 mM phosphate buffer pH 6 according to Thomassen et al.[5] Monosaccharide composition analysis was performed in triplicates by a modified NREL sulphuric acid hydrolysis.[6] In brief, 150 mg sample was added to 1.5 mL 72 % sulfuric acid, followed by incubation at 30 °C for 60 min. Acid concentration was diluted to a final 4 % sulfuric acid and the sample was autoclaved at 121 °C for 60 min. Standards were treated in the same manner. Quantification was performed by High-Performance Anion-Exchange Chromatography

Coupled with Pulsed Amperometric Detection (HPAEC-PAD) using a Dionex ICS-5000 system (DionexCorp, Sunnyvale, CA) equipped with a CarboPac PA1 analytical column (4x250 mm) and a guard column (4x50 mm), operated at 1 mL/min with isocratic elution in water for 38 min followed by isocratic elution in 500 mM NaOH for 15 min. Post-column addition of 500 mM NaOH at 0.2 mL/min was added for detection.

**Measurement of anthropometrics and blood pressure**

Subjects were weighed to the nearest 0.05 kg (Lindell Tronic 8000, Digital Medical Scale, Copenhagen, Denmark) in light clothing and with an empty bladder. At the first examination subjects additionally had their height measured to the nearest 0.5 cm by a wall-mounted stadiometer (Hultaforse, Sweden). Body composition was measured by bioelectrical impedance analysis (QuadScan 4000, Bodystat Inc, Isle of Man, British Isles, United Kingdom) and fat-free mass was calculated by subtracting fat mass from body weight. Waist circumference was measured twice to the nearest 0.5 cm using a flexible measuring tape (Meterex, Lagenfeld, Germany) at the point of the navel after an exhalation. Sagittal abdominal diameter (SAD) at an exhale was measured twice to the nearest 0.1 cm using an abdominal caliber (Holtain-Kahn Abdominal Caliper, Crosswell, United Kingdom) at the point of the navel with subjects lying on a flat bed with their legs bent. Systolic and diastolic blood pressure was measured with an automatic sphygmomanometer (A&D Medical, Tokyo, Japan) after the subject had been lying down for a least 10 min. Three assessments were made and the mean value of these was used in the statistical analysis.

**Biochemical analyses of blood samples**

All blood sample analyses were performed at the end of the study to ensure low variability. Plasma glucose, whole blood glycated hemoglobin (HbA1c) and serum total-, LDL- and HDL-cholesterol, triacylglycerol (TAG), free-fatty acids (FFA), alanine aminotransferase (ALT), and aspartate aminotransferase (AST) were analysed using automated, enzymatic, colorimetric assay on ABX

Pentra 400 chemistry analyzer (ABX Pentra, Horiba ABX, Montpellier, France). The coefficient of variance (CV) for these analyses was between 1.3-7.2 %. Analysis of serum insulin and C-peptide was measured by a chemiluminescent immunometric assay (Immulite 1000; Siemens Medical Solutions Diagnostics, Los Angeles, USA). The CV% was <5 % for both. Homeostatic model assessment for insulin resistance (HOMA-IR) was calculated according to Wallace et al[7] as insulin resistance = glucose * insulin / 135, where glucose is in mmol/L, and insulin is in pmol/L. Serum was diluted 1000 times, and C-reactive protein (CRP) was measured in a high-sensitivity single-plex assay (MesoScale Discovery®, Gaithersburg, MD, USA) using the Sector Imager 2400A (MesoScale Discovery®). The lower limit of detection was 4.3 pg/mL. Serum inflammatory markers IL-6, IL1β and TNF-α were measured using high-sensitivity enzyme linked immunosorbent assay (ELISA) (HS600B and HSTA00D, R&D systems, Minneapolis, Minnesota, USA), with CV% of 3.6-5.2 %. Plasma leptin was analysed by a quantitative enzyme immunoassay using a Microplate Reader (Tecan, Spectra III). Radioimmunological determinations of total plasma glucagon-like peptide-1 (GLP-1) and glucagon-like peptide-2 (GLP-2) concentrations were performed as described previously[8,9] with sensitivity <5 pmol/L and intra assay CV% <10 %.

Plasma alkylresorcinol homologs (C17:0, C19:0, C20:0, C21:0, C22:0, C23:0, C24:0, C25:0, and C26:0) were measured using a normal-phase liquid chromatography-tandem mass spectrometry method as described elsewhere.[10] Total plasma alkylresorcinol concentration was calculated as the sum of homologs C17:0-C26:0, and the C17:0-to-C21:0 ratio was calculated for each sample and used for statistical analyses. The concentrations of SCFAs in plasma were determined using a gas chromatography-mass spectrometry (GC-MS) method based on the method described by Brighenti.[11] In brief, plasma (400 μL) was mixed thoroughly with 100 μL internal standard solution containing 150 μM of acrylic acid (internal standard) and 14.4% (w/w) meta-phosphoric acid. The mixture was centrifuged in an Eppendorf centrifuge at 20817×g for 30 min and stored in at 4°C for 30 min, the supernatant (200 μL) was carefully transferred into a new test tube and extracted with propyl formate (200 μL) by vortex mixing for 5 min and then centrifuged at 20817×g

for 10 min. Propyl formate extract (upper phase) was collected (100 μL) in a GC vials and injected (2 μL) on a Finnigan Trace GC coupled to a Finnigan Trace DSQ II mass detector (Thermo Fisher Scientific, Waltham, MA, USA) with a ZB-FFAP column (7HG-G009-11. 30m × 0.25 mm ID and 0.25 μm film thickness) installed. Helium (2.0 mL min-1) was used as carrier gas. The inlet was equipped with an unpacked, straight glass liner and was set at 200°C in splitless mode. The oven temperature was held at 55°C for 4 min, then raised to 130°C within 1.5 min and held for 3.7 min, and finally raised to 250°C within 4 min and held for 2 min. Electron impact ionization (70eV) was applied at 250°C. Selected ion monitoring was carried out by monitoring the quantification ion (in bold) and confirmation ions together: m/z 60 and 43 for acetic acid, m/z 74 and 57 for propionic acid, m/z 73 and 60 for butyric acid, and m/z 72 and 60 for acrylic acid (internal standard). Quantification was made using a multipoint standard curve (5-640 μM for acetic acid, 0.5-64 μM for propionic acid and butyric acid, n=8). Known SCFAs concentrations were linearly regressed against the ratio of SCFAs/acrylic acid.

**Faecal microbiota profiling by metagenomics sequencing**

*Metagenomic sequencing*

DNA shearing and library preparations were performed according to the NEXTflex Rapid DNA-Seq Kit, V13.08 (Bioo Scientific, Austin, TX, USA). Briefly, 250 ng genomic DNA was sheared by Covaris E210 System using 10% duty cycle, intensity of 5, cycles per burst of 200 for 300 sec. To create 200bp fragments. The samples were end-repair and adenylated to produce an A-overhang. Adapters containing unique barcodes were ligated on to the DNA. The samples were then purified using the beads size selection for a selection range around 300-400bp with the Agercount AMPure XP beads (Beckman Culter, Beverly, MA, USA). The purified DNA libraries were amplified with a denaturation time of 2 minutes at 98°C, followed by 12 cycles of denaturation at 98°C for 30 seconds, annealing at 65°C for 30 seconds and extension at 72°C for 1 minute according to the protocol. The final extension was performed at 72°C for 4 minutes. Amplification was followed by

DNA quantification using NanoDrop ND- 1000 UV-VIS Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and the quality was checked on an Agilent 2100 Bioanalyzer using the Bioanalyser DNA High sensitivity (Agilent Technologies). Library preparation was performed by the DTU Multi-Assay Core (DMAC), Technical University of Denmark. The DNA libraries were mixed in equimolar ratios. Sequencing was performed as a 100 bp Pair-end run on HiSeq 2000 (Illumina Int., San Diego, CA USA) at BGI´s facility in Copenhagen following the manufacturer's recommendations. A complete set of successfully sequenced samples were available for 43 of the completers.

***Metagenomic sequence analysis***

The raw reads output was quality controlled using Cutadapt v. 1.8.1 with Python 2.7.10.[12] Bases with a Phred quality score below 20 were trimmed in both the 5' and 3' end of the sequences and reads shorter than 50 bases were discarded. Adaptor sequence remnants were removed in the 3' and 5' end of the reads. The sample read counts ranged from 730 Mbp to 16.8 Gbp. Trimmed reads were mapped against the Integrated Gene Catalogue (IGC)[13] comprising 9,879,896 genes derived from 1267 faecal samples, using BWA-MEM.[14] All samples reached the stationary plateau when rarefied to the catalogue. The read alignments were filtered using a 95% sequence similarity cutoff and the number of reads mapping to a gene was counted. Samples were rarefied using the R-package GUniFrac v 1.0 to the lowest sample size per individual to enable comparison within individual with a minimum loss of information. To infer the gene abundance, the rarefied gene counts were normalized by gene length.

The canopy clustering method described by Nielsen *et al*.[15] was applied on the 1267 samples from Li *et al*.[13] to create co-abundance gene clusters. Clusters containing more than 700 genes were assigned as metagenomic species (MGS). Abundances of genes belonging to such a cluster were used to calculate the overall relative abundance of an MGS. At least 5% of all genes belonging to an MGS should be observed before the MGS was detected; hence the $95^{th}$ quantile of the

abundances of the collection of genes from the same MGS was used as abundance measure. However, a minimum abundance cutoff was set to a $95^{th}$ quantile of the read count to genes normalized by gene length (in base pairs) of less than 1e-4.

To determine the taxonomy of the MGS, the genes were aligned using Blast[16] against a collected database create by concatenating the following databases: Bacteria (23 Feb 2016), Archaea (24 Nov 2015), Fungi (24 Nov 2015), Non Flu Viruses (2 Sep 2015), Plasmid (23 Feb 2016), Protozoa (24 Nov 2015) and Virus (24 Nov 2015), which were downloaded as described by Nordahl *et al*.[17] Blast hits with an e-value lower than 1e-5 and an alignment length above 80% of the query length were approved. The taxonomic annotation was based on the consistency of the taxonomy of approved blast hits by manual evaluation and validated using the 16S data by correlating MGSs with operational taxonomic units (OTUs) and comparing the taxonomic annotations.

To assign functions to genes, all IGC genes were annotated to CAZy (July 2016)[18] by sequence alignment,[16] where the best hit on alignment length and an e-value cutoff of 1e-5 was retained. Additionally, all IGC genes were annotated to KEGG (June 2016)[19] with a 50 alignment identity over 50 % alignment length cutoff. The best hit with lowest e-value (e-value cutoff 1e-5) was retained. Gene abundances of genes annotated to specific KO were summed up to KO bins per individual.

Sequencing data have been deposited to SRA database with the accession number PRJNA395744.

**Gut permeability assessment**

The percentage of excreted lactulose and mannitol in urine was measured by high-performance anion-exchange chromatography with electrochemical detection. Lactulose and mannitol were separated on a Dionex CarboPac MA1 BioLC Analytical 4x250 mm column, with 4x50 mm pre-column, with a mobile phase of 50 mM NaOH at a flow rate of 1 mL/min. Mannitol was also quantified by spectrophotometric analysis on a ABX Pentra 400 (Horiba Medical, California, USA)

(CV<10 %).

**Urine metabolomics profiling with GC-MS**

*Chemicals and reagents*

All chemicals used in the study were analytical grade. Sodium hydroxide, pyridine, methyl chloroformate (MCF), anhydrous sodium sulfate, sodium bicarbonate, the internal standard 2,3,3,3-$d_4$-DL-alanine and the external C7-C30 saturated alkanes standard were obtained from Sigma-Aldrich (St. Louis, MO, USA). Chloroform, methanol and hexane were obtained from Merck (Darmstadt, Germany).

*Quality control and standards*

Urine samples were randomised by a computer generated list of random numbers. However, all samples from the same subjects were analysed in the same daily batch to lower analytical variations within each subject. To ensure a high quality and reproducibility, several controls were included during the GC-MS analysis. Briefly, a batch of 20 urine samples was analysed daily by GC-MS together with chloroform blanks, a derivatised negative control containing $d_4$-alanine, a derivatised standard mixture containing 12 biologically relevant compounds (amino and non-amino organic acids), an external C7-C30 saturated alkanes standard (1 µg/µl diluted in hexane), and a derivatised quality control (QC) urine sample, which was injected for every 10 samples. While the blank control verified the absence of carry-over effects and the negative control verified the absence of contamination of reagents, the standard mixture and QC samples were included to evaluate the reproducibility of the derivatisation preparation from day to day. Finally, to monitor the stability of the GC-MS system throughout the experiment, a C7-C30 saturated alkanes standard, which did not undergo any derivatisation preparation, was injected daily.

*Derivatisation of samples*

Urine samples were thawed on ice. Prior to derivatisation, 100 µl of each urine sample was mixed with 80 µl of NaOH (3M) in a silanised reaction glass tube and 20 µl internal standard $d_4$-alanine (10 mM) to compensate for the technical variability. The standard mixtures and QC samples were prepared in the same way. A negative control was prepared by mixing 20 µl $d_4$-alanine with 180 µl NaOH (3M) in a silanised reaction glass tube. Next, the negative control, urine samples, standard mixtures and QC samples were derivatised according to the MCF protocol described by Smart *et al*.[20]

*GC-MS Analysis*

The derivatised samples were on the same day of preparation injected on a GC-MS (Agilent 7890A coupled to MSD 5975C) with a mass selective detector (EI) operating at 70 eV. The column used for all analyses was a TG-1701MS GC column (30m x 0.255mm x 0.1µm with 5m guard column, Thermo Fisher, USA). The GC-MS variables were set according to Smart *et al*.[20]

*Raw data processing and identification of metabolites*

Raw GC-MS data files were initially examined by MSD ChemStation and exported to AIA format (cdf-files). AMDIS software (NIST, Boulder, CO, USA) was subsequently used for deconvoluting GC-MS chromatograms and identifying metabolites using our in-house MCF MS library, which contains information of mass fragmentation and retention times of more than 300 authentic standards. The commercially available library from National Institute of Standards and Technology (NIST) was applied to assist the identification of metabolites not identified by our in-house library. We followed the metabolite identification protocols outlined by the metabolomics standard initiative (MSI)[21] and identified 81 metabolites by our in-house library (MSI level 1), 6 putatively annotated metabolites by searching the NIST library (MSI level 2), and 9 unknown metabolites (MSI level 4). A high confidence match was ensured by a minimum match factor of 75%[20] when searching against our in-house library and a mass-spectral match greater than 90% when searching

against the NIST library. The GC-MS data processed by AMDIS were then analysed by an in-house R package, which is an update from the *Metab* R package[22] that makes use of the XCMS functions[23] to generate the raw GC-MS profiles. The GC-MS data (peak height) were normalised by the internal standard $d_4$-alanine and the measured creatinine concentration to account for the technical variability associated with chemical derivatisation and the dilution of urine. The coefficient of variation (CV) was calculated for all metabolites detected in the QC samples. 20 metabolites with CV > 30% were excluded as these were considered to have poor reproducibility[24].

**Metabolite identification by UPLC-MS**

The accurate masses of the discriminating features measured by UPLC-MS were searched for putative molecules that matched the METLIN[25] and HMDB[26] databases. For the identification of UPLC-MS metabolite candidates, discriminating features were subjected to MS/MS experiments. The metabolites were identified according to the four different levels described by the MSI;[21] metabolites confirmed by an authentic standard (Level I), metabolites confirmed based on a comparison of MS/MS fragmentation pattern compared with those found in databases and earlier literature (Level II), metabolites with similarities to published fragmentation patterns (Level III), and unknown compounds (Level IV). Sulfonated and glucuronidated metabolites were identified by deconjugation experiments using β-glucuronidase from *E. coli* K12 (Roche Diagnostics GmbH) and sulfatase from *Aerobacter aerogenes* (Sigma-Aldrich, Schnelldorf, Germany). For each deconjugation reaction, urine samples were diluted in sodium phosphate buffer (pH=7.4) and incubated with the enzymes at 37°C for 1-2 hours and injected onto the UPLC-MS for analysis and compared to standards where available. Authentic standards were obtained for the following metabolites: 3-(3,5-dihydroxyphenyl)-1-propanoic acid (DHPPA), 2-aminophenol, and pyrocatechol purchased from Sigma-Aldrich (Schnelldorf, Germany).

**Sample size estimation**

We based our power calculation for HOMA-IR on effects observed in previous studies,[27–29] both with respect to SD and effect size (0.25 units), since we aimed to either verify or disprove these effects and address the current inconsistency about whether or not wholegrain has an effect on glucose metabolism. The applied effect size of HOMA-IR (0.25 units) is equivalent to approximately 1/6 of the currently reported effects of intensive lifestyle interventions and metformin treatment (1.5-2.0 units), both of which have been shown to be prevent type 2 diabetes in individuals with impaired glucose tolerance.[30,31] However, the subjects included in this study did not all have impaired fasting glucose, and the overall effects size were therefore expected to be lower.

## SUPPLEMENTARY FIGURES

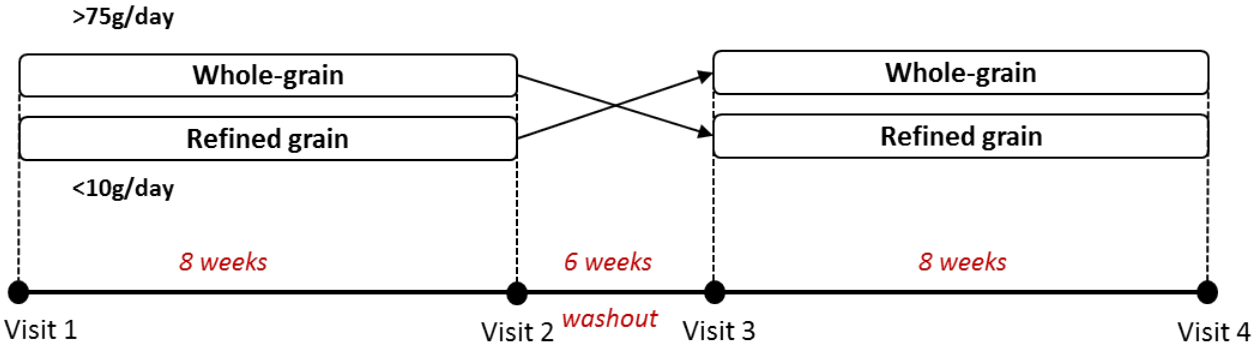


**Figure S1.** Study outline. The study applied a randomized, controlled cross-over design comprising two dietary intervention periods of eight weeks duration, separated by a wash-out period of at least six weeks. We compared the effects of the dietary interventions on the primary and secondary outcomes of the trial using measurements taken at baseline (visit 1), post-randomization baseline (visit 3) and end-points (visit 2 and visit 4). This was done by using a linear mixed model adjusting for age, gender, carry-over effects (i.e. the putative effect of the treatment from the first intervention period on the post-randomization baseline) and period effects (i.e. the effect of time which occurred independent of the given treatment)

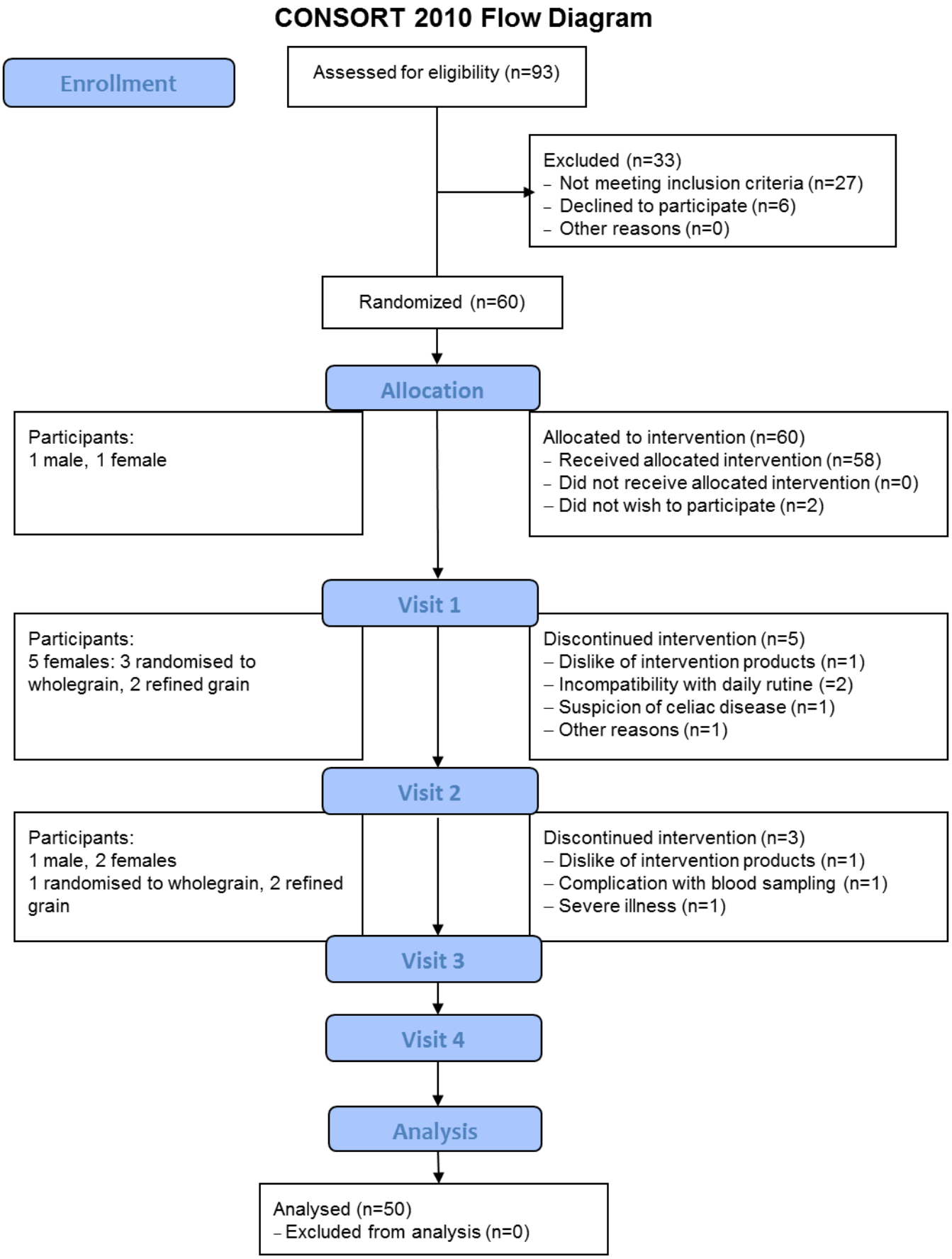


**Figure S2.** CONSORT flow diagram for the study.

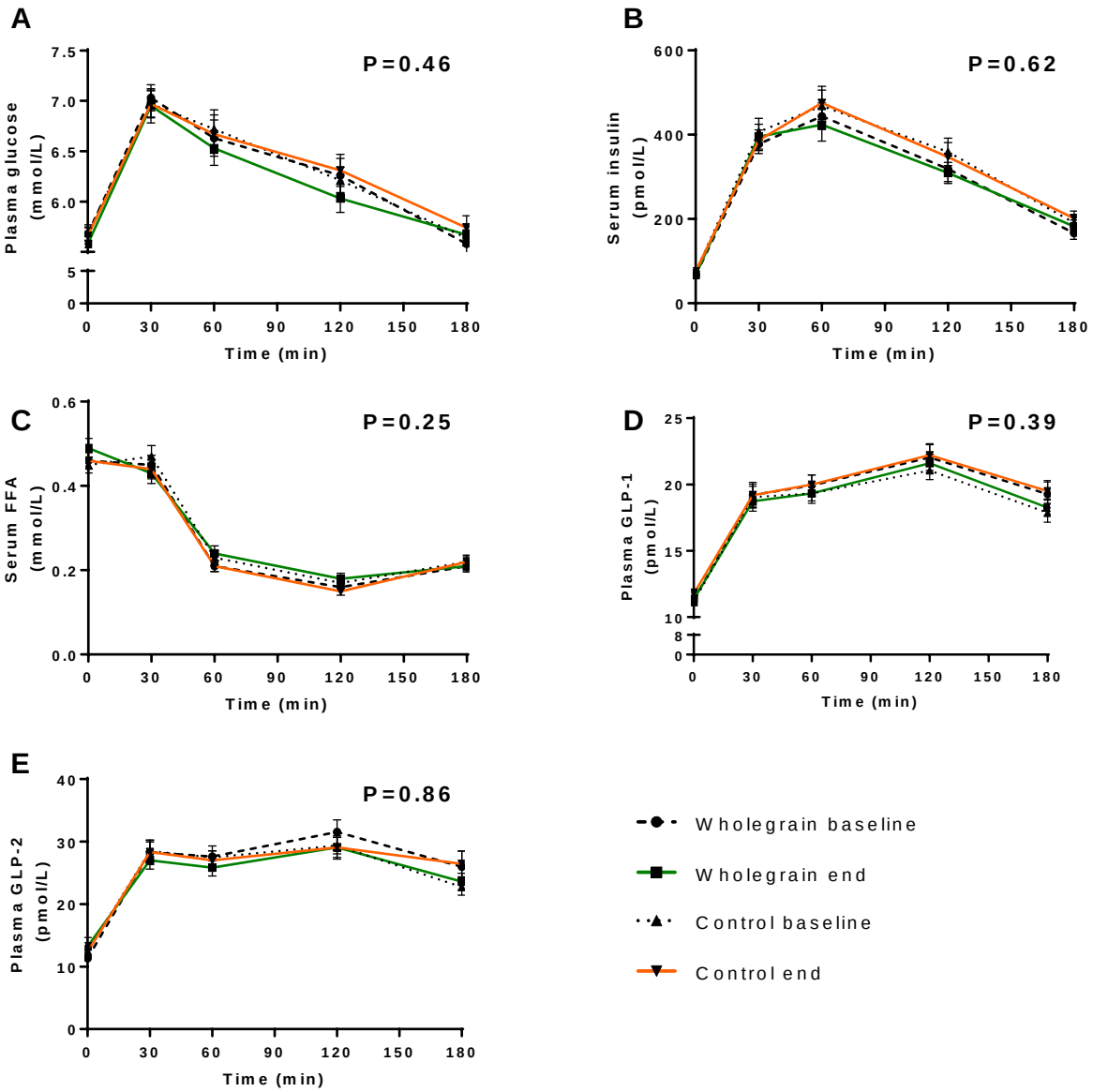


**Figure S3.** Postprandial responses after a standardized, non-wholegrain breakfast (including a lactulose/mannitol drink) before and after the 8-week intervention periods with refined grain and wholegrain diets. Data are mean±SEM (n=50). Fasting (t=0) and postprandial concentrations of (**A**) plasma glucose, (**B)** serum insulin, (**C**) serum free fatty acid (FFA), (**D**) plasma glucagon-like peptide 1 (GLP-1), and (**E**) plasma glucagon-like peptide 2 (GLP-2) were not affected by refined grain or wholegrain diet as assessed by the linear mixed model adjusting for age and gender.

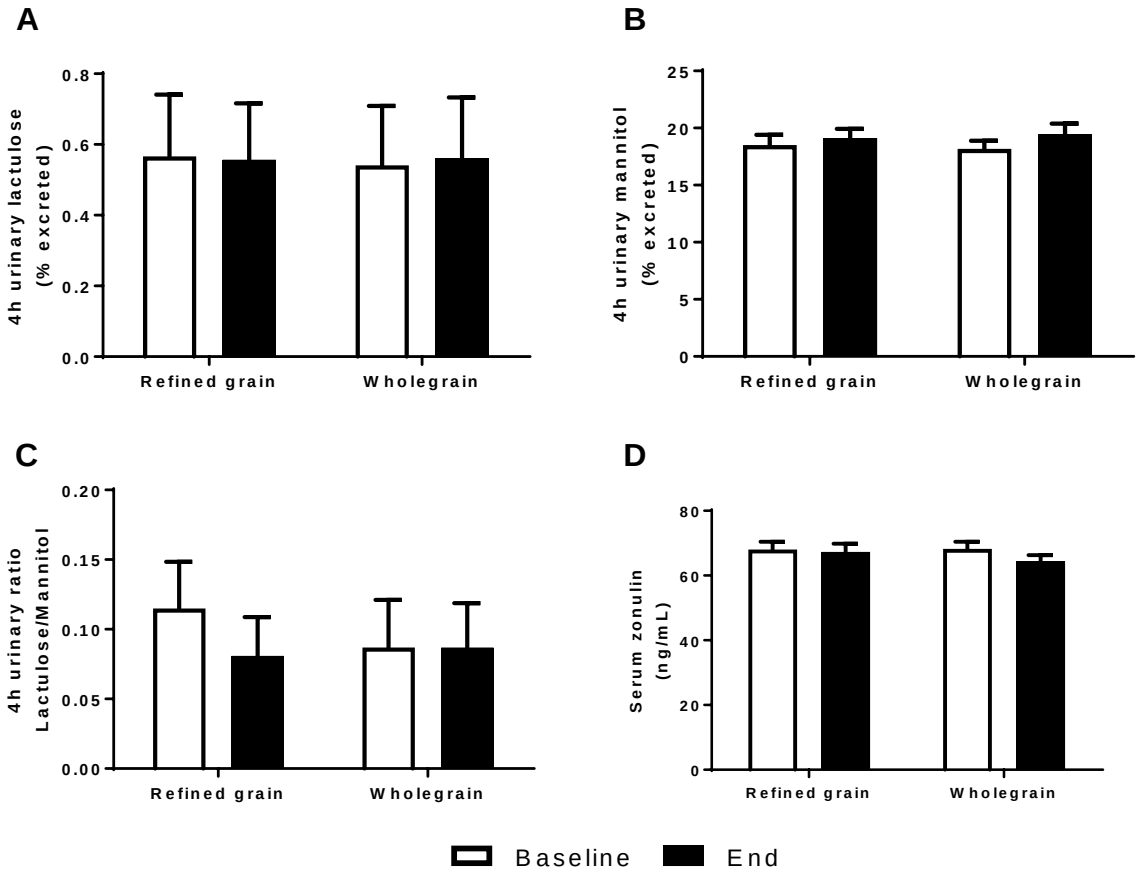


**Figure S4** Biomarkers of gut permeability after consumption of the refined grain and the wholegrain diet for 8 weeks each. Bars represent means ± SEM (n=50). Gut permeability expressed by the urinary excretion of (**A**) lactulose and (**B**) mannitol and (**C**) the lactulose/mannitol ratio, and by (**D**) serum concentrations of zonulin were not significantly different between the two dietary interventions as assessed by the linear mixed model adjusting for age and gender.

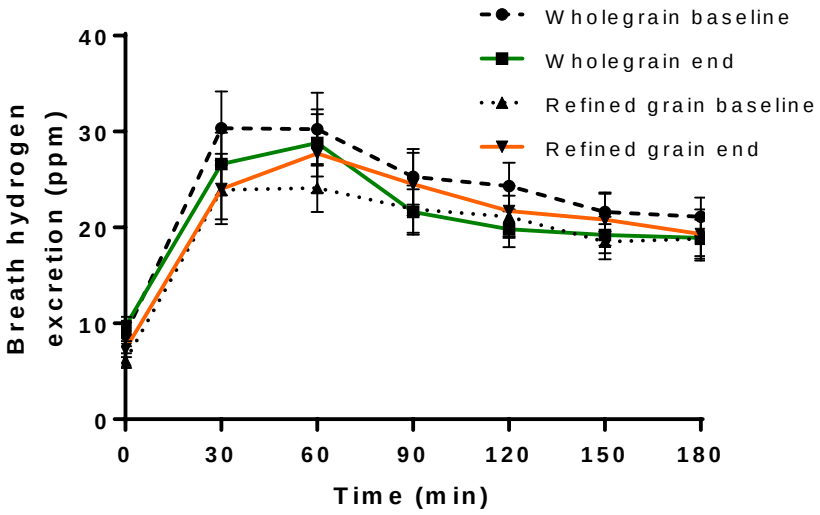


**Figure S5.** Breath hydrogen excretion after a standardized, non-wholegrain breakfast (including a lactulose/mannitol drink) before and after the refined grain and wholegrain diet. Data are mean ± SEM (n=50). Ppm; parts per million. Breath hydrogen was not significantly different between refined grain diet and wholegrain diet as assessed by linear mixed model adjusting for age and gender.

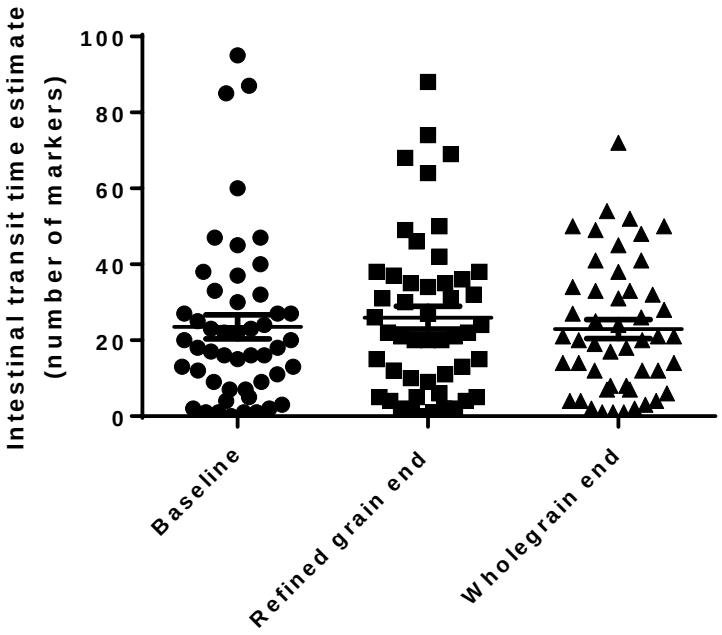


**Figure S6**. Intestinal transit time estimates at baseline and after the dietary interventions with wholegrain and refined grain. Intestinal transit time was estimated based on number of radiopaque markers visible on abdominal X-ray and adjusted for time since last marker ingestion. Data are mean ± SEM (n=50). No significant differences in intestinal transit time were found as assessed by one-way repeated measures ANOVA followed by Tukey's Multiple Comparison test.

## SUPPLEMENTARY REFERENCES


1 Ibrügger S, Vigsnæs LK, Blennow A, *et al.* Second meal effect on appetite and fermentation of wholegrain rye foods. *Appetite* 2014;**80**:248–56. doi:10.1016/j.appet.2014.05.026

2 Akerberg AK, Liljeberg HG, Granfeldt YE, *et al.* An in vitro method, based on chewing, to predict resistant starch content in foods allows parallel determination of potentially available starch and dietary fiber. *J Nutr* 1998;**128**:651–60.http://www.ncbi.nlm.nih.gov/pubmed/9482777.

3 Courtin CM, Van den Broeck H, Delcour JA. Determination of reducing end sugar residues in oligo- and polysaccharides by gas-liquid chromatography. *J Chromatogr A* 2000;**866**:97–104.http://www.ncbi.nlm.nih.gov/pubmed/10681013.

4 Ross AB. Analysis of alkylresorcinols in cereal grains and products using ultrahigh-pressure liquid chromatography with fluorescence, ultraviolet, and CoulArray electrochemical detection. *J Agric Food Chem* 2012;**60**:8954–62. doi:10.1021/jf301332q

5 Thomassen L V., Meyer AS. Statistically designed optimisation of enzyme catalysed starch removal from potato pulp. *Enzyme Microb Technol* 2010;**46**:297–303. doi:10.1016/j.enzmictec.2009.11.003

6 Sluiter A, Hames B, Ruiz R, *et al.* Determination of structural carbohydrates and lignin in biomass. *Lab Anal Proced* 2008;**1617**.

7 Wallace TM, Levy JC, Matthews DR. Use and abuse of HOMA modeling. *Diabetes Care* 2004;**27**:1487–95.http://www.ncbi.nlm.nih.gov/pubmed/15161807.

8 Færch K, Torekov SS, Vistisen D, *et al.* GLP-1 Response to Oral Glucose Is Reduced in Prediabetes, Screen-Detected Type 2 Diabetes, and Obesity and Influenced by Sex: The ADDITION-PRO Study. *Diabetes* 2015;**64**:2513–25. doi:10.2337/db14-1751

9 Hartmann B, Johnsen AH, Orskov C, *et al.* Structure, measurement, and secretion of human glucagon-like peptide-2. *Peptides* 2000;**21**:73–80.http://www.ncbi.nlm.nih.gov/pubmed/10704722.

10 Ross AB, Svelander C, Savolainen OI, *et al.* A high-throughput method for liquid chromatography-tandem mass spectrometry determination of plasma alkylresorcinols, biomarkers of whole grain wheat and rye intake. *Anal Biochem* 2016;**499**:1–7. doi:10.1016/j.ab.2015.12.023

11 Brighenti F, Gullion F, Amado R, *et al.* Functional Properties of Non-digestible Carbohydrates. Brussels: European Commission, DG XII. In: *Summary of the conclusion of the working group on Profibre interlaboratory study on determination of short chain fatty acids in blood*. 1998. 150–3.

12 Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal; Vol 17, No 1 Next Gener Seq Data Anal* Published Online First: 2011.http://journal.embnet.org/index.php/embnetjournal/article/view/200

13 Li J, Jia H, Cai X, *et al.* An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* 2014;**32**:834–41. doi:10.1038/nbt.2942

14 Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;**26**:589–95. doi:10.1093/bioinformatics/btp698

15 Nielsen HB, Almeida M, Juncker AS, *et al.* Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat Biotechnol* 2014;**32**:822–8. doi:10.1038/nbt.2939

16 Altschul SF, Gish W, Miller W, *et al.* Basic local alignment search tool. *J Mol Biol* 1990;**215**:403–10.

17 Nordahl Petersen T, Rasmussen S, Hasman H, *et al.* Meta-genomic analysis of toilet waste from long distance flights; a step towards global surveillance of infectious diseases and antimicrobial resistance. *Sci Rep* 2015;**5**:11444. doi:10.1038/srep11444

18 Lombard V, Golaconda Ramulu H, Drula E, *et al.* The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 2014;**42**:D490-5. doi:10.1093/nar/gkt1178
19 Kanehisa M, Sato Y, Kawashima M, *et al.* KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 2016;**44**:D457-62. doi:10.1093/nar/gkv1070
20 Smart KF, Aggio RBM, Houtte JVR Van, *et al.* Analytical platform for metabolome analysis of microbial cells using methyl chloroformate derivatization followed by gas chromatography-mass spectrometry. *Nat Protoc* 2010;**5**:1709–29. doi:10.1038/nprot.2010.108
21 Sumner LW, Amberg A, Barrett D, *et al.* Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* 2007;**3**:211–21. doi:10.1007/s11306-007-0082-2
22 Aggio R, Villas-Bôas SG, Ruggiero K. Metab: an R package for high-throughput analysis of metabolomics data generated by GC-MS. *Bioinformatics* 2011;**27**:2316–8. doi:10.1093/bioinformatics/btr379
23 Smith CA, Want EJ, O'Maille G, *et al.* XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal Chem* 2006;**78**:779–87.http://dx.doi.org/10.1021/ac051437y
24 Dunn WB, Broadhurst D, Begley P, *et al.* Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat Protoc* 2011;**6**:1060–83. doi:10.1038/nprot.2011.335
25 Smith C, O'Maille G, Want E, *et al.* METLIN: a metabolite mass spectral database. *Ther Drug Monit* 2005;**27**:747–51.
26 Wishart DS, Jewison T, Guo AC, *et al.* HMDB 3.0--The Human Metabolome Database in 2013. *Nucleic Acids Res* 2013;**41**:D801–7. doi:10.1093/nar/gks1065
27 Pereira MA, Jacobs DR, Pins JJ, *et al.* Effect of whole grains on insulin sensitivity in overweight hyperinsulinemic adults. *Am J Clin Nutr* 2002;**75**:848–55.http://www.ncbi.nlm.nih.gov/pubmed/11976158.
28 Kristensen M, Toubro S, Jensen MG, *et al.* Whole grain compared with refined wheat decreases the percentage of body fat following a 12-week, energy-restricted dietary intervention in postmenopausal women. *J Nutr* 2012;**142**:710–6. doi:10.3945/jn.111.142315
29 Tighe P, Duthie G, Vaughan N, *et al.* Effect of increased consumption of whole-grain foods on blood pressure and other cardiovascular risk markers in healthy middle-aged persons: a randomized controlled trial. *Am J Clin Nutr* 2010;**92**:733–40. doi:10.3945/ajcn.2010.29417
30 Knowler WC, Barrett-Connor E, Fowler SE, *et al.* Reduction in the Incidence of Type 2 Diabetes with Lifestyle Intervention or Metformin. *N Engl J Med* 2002;**346**:393–403. doi:10.1056/NEJMoa012512
31 Haffner S, Temprosa M, Crandall J, *et al.* Intensive lifestyle intervention or metformin on inflammation and coagulation in participants with impaired glucose tolerance. *Diabetes* 2005;**54**:1566–72.http://www.ncbi.nlm.nih.gov/pubmed/15855347.