SUPPLEMENTARY FILE 1

Mice and diets

In vivo studies were performed in accordance with European guidelines for the use and care of laboratory animals and approved by an independent ethics committee (authorization 17430-2018110611093660 v3).

Thirteen-week-old male and female C57BL/6J mice (purchased from Charles River, n=12 per group and per sex) were fed a chow diet (CTRL, D12450J, Research Diets), a high-fat diet (HFD, D12492, Research Diets), choline-deficient high-fat diet (CDHFD, D05010402, Research Diets), Western diet (WD, TD.88137, Envigo), or WD with glucose (18.9 g/L) and fructose (23.1 g/L) in drinking water (WD GF)[1] for 15 weeks. Body weight and food intake were measured weekly. Mice were killed at ZT16.

Control diet contains 20% proteins, 70% carbohydrates, 10% fat (3.85 kcal/g), HFD contains 20% proteins, 20% carbohydrates, 60% fat (5.24 kcal/g), CD-HFD contains 20% proteins, 35.1% carbohydrates, 44.9% fat (4.74 kcal/g), WD contains 15.2% proteins, 42.7% carbohydrates, 42% fat, 0.2% cholesterol (4.5 kcal/g) co administrated or not with glucose (18.9 g/L) and fructose (23.1 g/L) in drinking water.

Pparα$^{hep-/-}$ mice

Pparα$^{hep-/-}$ animals were created at INRA's rodent facility (Toulouse, France) by mating the floxed-*Pparα* mouse strain with C57BL/6J albumin-Cre transgenic mice (a gift from Prof. Didier Trono, EPFL, Lausanne, Switzerland) to obtain albumin-Cre+/−*Pparα*$^{flox/flox}$ mice (i.e., *Pparα*$^{hep-/-}$ mice). The *Pparα* deletion was confirmed with PCR and HotStar Taq DNA Polymerase (5 U/µl, Qiagen) using the following primers: forward: 5′-AAAGCAGCCAGCTCTGTGTTGAGC-3′ and reverse, 5′-TAGGTACCGTGGACTCAGAGCTAG-3′. The amplification conditions were as follows: 95°C for 15 min, followed by 35 cycles of 94°C for min, 65°C for 1 min, and 72°C for 1 min, and a final cycle of 72°C for 10 min. This reaction produced 450-bp, 915-bp, and 1070-bp fragments, which represented the *Pparα* sequence with an exon 4 deletion, the wild-type allele, and the floxed allele, respectively. The albumin-Cre allele was detected by PCR using the following primer pairs: CreU, 5′-AGGTGTAGAGAAGGCACTTAG-3′ and CreD, 5′- CTAATCGCCATCTTCCAGCAGG-

3′; G2lox7F, 5′-CCAATCCCTTGGTTCATGGTTGC-3′ and G2lox7R, 5′-CGTAAGGCCCAAGGAAGTCCTGC-3′).

Western diet experiments with *Pparα*$^{hep-/-}$ mice

Thirteen-week old male and female *Pparα*$^{hep+/+}$ and *Pparα*$^{hep-/-}$ mice were fed a standard diet (Safe A04) or a Western Diet (WD, TD.88137, Envigo) during 15 weeks.

Fasting experiments with *Pparα*$^{hep-/-}$ mice

Fourteen-week old male and female *Pparα*$^{hep+/+}$ and *Pparα*$^{hep-/-}$ mice (n=12 per group) were fasted for 24h (starting at ZT0) to measure kinetics of ketone bodies production and then fed ad libitum. Two weeks later, these mice were either fasted for 16h (starting at ZT0) or fed *ad libitum* with a standard diet (Safe A04), (n=6 per group). Mice were sacrificed at ZT16.

Pemafibrate experiments with *Pparα*$^{hep-/-}$ mice

Ten-week old male and female *Pparα*$^{hep+/+}$ and *Pparα*$^{hep-/-}$ mice (n=8 per group) fed a standard diet (Safe A04) received 0.1 mg/kg/day pemafibrate (Kowa, MedChemExpress) or vehicle (carboxymethylcellulose 0.5%) during 14 days. Gavages were performed every day at ZT16. Mice were sacrificed at ZT16, fed ad libitum, the last gavage was performed the day before the sacrifice.

Oral glucose tolerance test (OGTT)

OGTT was performed after 13 weeks of diet. Mice were fasted for 6 h and received an oral glucose load (2g/kg body weight). Blood glucose was measured at the tail vein using an AccuCheck Performa glucometer (Roche Diagnostics) at -15, 0, 15, 30, 45, 60, 90, and 120 minutes.

Blood and tissue sampling in mice

Prior to sacrifice, blood was collected into EDTA coated tubes (BD Microtainer, K2E tubes) from the submandibular vein. All mice were killed in a fed state. Plasma was collected by centrifugation (1500 xg, 10min, 4°C) and stored at -80°C. Following killing by cervical dissociation, organs were removed, weighted, dissected and used for histological analysis or snap frozen in liquid nitrogen and stored at -80°C.

Plasma Insulin and FGF21 were assayed using the rat/mouse Insulin and FGF21 ELISA kit (Sigma) according to the manufacturer's instructions.

Mouse adiponectin and leptin concentrations were determined at the Anexplo-Crefre Phenotyping facility, using the xMAP Technology (Magpix), with a R&D System Kit.

Free carnitine and acylcarnitines were measured from plasma (10 µL) spotted on filter membranes (Protein Saver 903 cards; Whatman), dried, and then treated as reported.[2] Briefly, acylcarnitines were derivatized to their butyl esters and treated with the reagents of the NeoGram MSMS-AAAC kit (PerkinElmer). Their analysis was carried out on a Waters 2795/Quattro Micro AP liquid chromatography–tandem mass spectrometer (Waters, Milford, MA).

Human liver samples from patients with NAFLD

Liver samples were selected from a cohort of biopsy-proven NAFLD patients established in the Hepatology Department of the Centre Hospitalier Universitaire de Toulouse. All patients gave written consent and liver samples were frozen and stored at - 80°C at the Biological Resource Center of the Toulouse University Hospital. For the present study, patients who met the following criteria were selected: age over 18, signed informed consent, liver biopsy puncture (PBH) performed by the transjugular route. Patients with one or more of the following criteria were excluded from the analysis: women hysterectomized before 50 years with unknown hormonal status, Turner's syndrome, hormone treatments other than contraceptive pill and hormone replacement therapy, immunosuppressive treatments, steatosic treatments (methotrexate, tamoxifen, corticosteroids, amiodarone), patient under guardianship or curatorship, and liver biopsy performed during bariatric surgery. Clinical history of menses cessation was systematically recorded for all women included in the cohort, as well as the use of hormonal treatments. No women included in the analysis was on hormonal contraceptives or menopause therapy. The 80 biopsies were from 48 men and 32 women (78% were post-menopausal). This cohort was approved by the National Agency for the Safety of Medicines and Health Products (ANSM) and local ethics committee in 2015 (ClinicalTrial.gov: NCT02390232), and by the Minister of Higher Education, Research, and Innovation in 2017 (DC-2017-2984).

Histological analysis

Paraformaldehyde-fixed, paraffin-embedded liver tissue was sliced into 3-µm sections and stained with haematoxylin and eosin (H&E). Steatosis and the inflammation (NAFLD activity score or NAS) for mouse liver were evaluated according to [3]. Steatosis was measured depending on the percentage of liver cells containing fat (Grade 0 to 3). The degree of inflammation was appreciated by counting the inflammatory foci into 10 distinct areas at 200X for each liver slice (Grade 0 to 3). Sirius red-stained sections were used for evaluation of fibrosis as follows: 0, no fibrosis; 1, pericellular and perivenular fibrosis; and 2, focal bridging fibrosis. Immunohistochemical staining of anti-a-smooth muscle actin (a-SMA) (Abcam®) was also performed. a-SMA labeling score was established according to Akpolat et al.[4] as follows; score 0: absence of labeling or <3% of the periportal region; 1: between 3 and 33%; 2: between 34 and 66% and 3:> 67% of the periportal region. The images were taken by Olympus microscope at 10x magnification. Morphometric pixel analysis was quantified on 3 non-overlapping fields per mouse by using Image J software (NIH, USA). Results are the mean value from these three images, which represented one mouse.

In human liver samples, NAFLD was defined by fatty liver found on biopsy after exclusion of other causes of steatosis (iatrogenic etiology, excessive alcohol consumption, chronic hepatitis B or C and other causes of chronic liver diseases discovered on the biopsy). Histological analysis of the liver samples was performed by a liver pathology expert (Pr Janick Selves, Oncopole, Toulouse, France). NAFLD was evaluated by the SAF (Steatosis, Activity, Fibrosis) score.[5] NASH was defined by the presence of the following three criteria: steatosis grade ≥ 1, lobular inflammation grade ≥ 1 and balloonization score ≥ 1.[5] Fibrosis was classified from F0 to F4: F0 = no fibrosis, F1 = central fibrosis or portal, F2 = central and portal fibrosis, F3 = bridging fibrosis, F4 = cirrhosis. Severe fibrosis was defined by an F score ≥ 3. NAFLD activity (NAS) (0-8) corresponded to the sum of the scores for steatosis (0-3), ballooning (0-2) and lobular inflammation (0-3).[3]

RNA extraction and RT-qPCR

RNAs were extracted from liver samples using the method with Tri-reagent [6] (Sigma Aldrich France). RNAs were quantified using nanodrop (ThermoScientific Nanodrop1000, Les Ulis, France). Two micrograms of total RNA were reverse-

transcribed using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). The SYBR Green assay primers are presented below.

Gene	NCBI Refseq	Forward primer	Reverse primer
Ccl2	NM_011333	GGTGTCCCAAAGAAGCTGTAGTTTT	AGTTGTAGGTTCTGATCTCATTTGGTT
Col1a1	NM_007742	GGCTCCTGCTCCTCTTAGGG	TCGGGTTTCCACGTCTCAC
Cxcl10	NM_021274	TGCCGTCATTTTCTGCCTC	GGCCCGTCATCGATATGG
Cyp4a10	NM_010011	TCCAGCAGTTCCCATCACCT	TTGCTTCCCCAGAACCATCT
Cyp4a14	NM_007822	TCAGTCTATTTCTGGTGCTGTTC	GAGCTCCTTGTCCTTCAGATGGT
Ehhadh	NM_023737	CGTCTCCTCGGTTGGTGTTC	ATTATCTTCTTTGCAGTATCTAGCTGCTT
Elovl3	NM_007703	GCCTCTCATCCTCTGGTCCT	TGCCATAAACTTCCACATCCT
Fgf21	NM_020013.4	AAAGCCTCTAGGTTTCTTTGCCA	CCTCAGGATCAAAGTGAGGCG
Cidec	NM_178373	CATGAAGTCTCTCAGCCTCCTGTA	CAGCTGTTGGGTCACCACTG
Mogat1	NM_026713	TGCCCTATCGGAAGCTGATCTA	CAGAGTCTGCTGAACAGGGATG
Pnpla3	NM_054088	ACGCGGTCACCTTCGTGT	AGCCCGTCTCTGATGCACTT
Pparα	NM_011144	CCCTGTTTGTGGCTGCTATAATTT	GGGAAGAGGAAGGTGTCATCTG
Pparg2	NM_011146	GATGCACTGCCTATGAGCACTT	GAATGGCATCTCTGTGTCAACC
Scd1	NM_009127	CAGTGCCGCGCATCTCTAT	CAGCGGTACTCACTGGCAGA
Tbp	NM_013684	ACTTCGTGCAAGAAATGCTGAA	GCAGTTGTCCGTGGCTCTCT
Tnf	NM_013693	TCCCCAAAGGGATGAGAAGTTC	GCGCTGGCTCAGCCACT
Vnn1	NM_011704	ATGAGGTTTATGCCTTTGGAGC	CCACAGGTGCGTAAATTGGTAG

Amplification was performed using an ABI Prism 7300 Real-Time PCR System (Applied Biosystems). qPCR data were normalized to TATA-box-binding protein mRNA levels and analyzed with LinRegPCR.v2015.3 to get the starting concentration (N0) which is calculated as follow: N0 = threshold/(Eff mean^{Cq}) with Eff mean: mean PCR efficiency and Cq: quantification cycle.

Microarray gene expression studies

Gene expression profiles were obtained for six liver samples per group in animal experiments and for 80 patients at the GeT-TRiX facility (GénoToul, Génopole Toulouse Midi-Pyrénées) using Agilent Sureprint G3 Mouse GE v2 microarrays (8x60K, design 074809) following the manufacturer's instructions.[7–9]

For each sample, Cyanine-3 (Cy3) labelled cRNA was prepared from 200 ng of total RNA (100 ng in human samples) using the One-Color Quick Amp Labeling kit (Agilent) according to the manufacturer's instructions, followed by Agencourt RNAClean XP (Agencourt Bioscience Corporation, Beverly, Massachusetts). Dye incorporation and cRNA yield were checked using a Dropsense™ 96 UV/VIS droplet reader (Trinean, Belgium). A total of 600 ng of Cy3-labelled cRNA was hybridized on the microarray slides following the manufacturer's instructions. Immediately after washing, the slides were scanned on an Agilent G2505C Microarray Scanner using Agilent Scan Control A.8.5.1 software and the fluorescence signal extracted using Agilent Feature Extraction software v10.10.1.1 with default parameters. Microarray data and experimental details are available in NCBI's Gene Expression Omnibus- [10] and are accessible through GEO Series accession number GSE159090 (diet experiment), GSE159089 (fasting experiment), GSE159086 (pemafibrate experiment), GSE159088 (human biopsies analysis).

Microarray data were analysed using R (www.r-project.org, R v. 3.1.2) and Bioconductor packages (www.bioconductor.org, v 3.0) as described in GSE159090 (diet experiment), GSE159089 (fasting experiment), GSE159086 (pemafibrate experiment), GSE159088 (human biopsies analysis).

Raw data (median signal intensity) were filtered, log2 transformed, and normalized using the quantile method.[11] A model was fitted using the limma lmFit function.[7] Pair-wise comparisons between biological conditions were applied using specific contrasts. A correction for multiple testing was applied using the Benjamini-Hochberg procedure [12] for false discovery rate (FDR). Hierarchical clustering was applied to the samples and the differentially expressed probes using 1-Pearson correlation coefficient as distance and Ward's criterion for agglomeration. The clustering results are illustrated as a heatmap of expression signals. Functional annotation clustering of KEGG pathways were performed using String database.[9]

Reporter metabolite analysis

Reporter metabolite analysis was performed to investigate the sex-dependent metabolites affected by transcriptional changes in response to diet using the PIANO package, which enriches the gene set analysis of genome-wide data by incorporating the directionality of gene expression and combining statistical hypotheses and methods, in conjunction with a previously established genome-scale metabolic model for the liver, iHepatocyte2322.[13]

Liver neutral lipids analysis

Hepatic lipid contents were determined at the end of the experiment as described elsewhere [14]. Briefly, following homogenization of tissue samples in 2:1 (v/v) methanol/EGTA (5 mM), lipids corresponding to an equivalent of 2 mg of tissue were extracted according to Bligh and Dyer in chloroform:methanol:water (2.5:2.5:2.1, v/v/v) in the presence of the internal standards glyceryltrinonadecanoate, stigmasterol, and cholesteryl heptadecanoate (Sigma, Saint-Quentin-Fallavier, France). Triglycerides, free cholesterol, and cholesterol esters were analysed by gas chromatography on a Focus Thermo Electron system using a Zebron-1 Phenomenex fused-silica capillary column (5 m, 0.32 mm i.d., 0.50 µm film thickness). The oven temperature was programmed from 200 to 350°C at a rate of 5°C/min, and the carrier gas was hydrogen (0.5 bar). The injector and detector were at 315°C and 345°C, respectively.

Liver fatty acid analysis

To measure all hepatic fatty acid methyl ester (FAME) molecular species, lipids corresponding to an equivalent of 2 mg of liver were extracted in the presence of the internal standard, glyceryl triheptadecanoate (2 µg). The lipid extract was transmethylated with 1 ml BF3 in methanol (14% solution; Sigma) and 1 ml heptane for 60 min at 80°C and evaporated to dryness. The FAMEs were extracted with heptane/water (2:1). The organic phase was evaporated to dryness and dissolved in 50 µl ethyl acetate. A sample (1 µl) of total FAMEs was analyzed with gas-liquid chromatography (Clarus 600 Perkin Elmer system, with Famewax RESTEK fused silica capillary columns, 30-m×0.32-mm i.d., 0.25-µm film thickness). Oven temperature was programmed to increase from 110°C to 220°C at a rate of 2°C/min,

and the carrier gas was hydrogen (7.25 psi). Injector and detector temperatures were 225°C and 245°C, respectively.

Liver phospholipid and sphingolipid analysis

The liquid chromatography solvent, acetonitrile, was HPLC-grade and purchased from Acros Organics. Ammonium formate (>99%) was supplied by Sigma Aldrich. Synthetic lipid standards (Cer d18:1/18:0, Cer d18:1/15:0, PE 12:0/12:0, PE 16:0/16:0, PC 13:0/13:0, PC 16:0/16:0, SM d18:1/18:0, SM d18:1/12:0) were purchased from Avanti Polar Lipids. Lipids were extracted from the liver (1 mg) as described by Bligh and Dyer in dichloromethane/methanol (2% acetic acid)/water (2.5:2.5:2 v/v/v). Internal standards were added (Cer d18:1/15:0, 16 ng; PE 12:0/12:0, 180 ng; PC 13:0/13:0, 16 ng; SM d18:1/12:0, 16 ng; PI 16:0/17:0, 30 ng; PS 12:0/12:0, 156.25 ng). The solution was centrifuged at 1500 rpm for 3 min. The organic phase was collected and dried under azote, then dissolved in 50 µl MeOH. Sample solutions were analysed with an Agilent 1290 UPLC system coupled to a G6460 triple quadripole spectrometer (Agilent Technologies). MassHunter software was used for data acquisition and analysis. A Kinetex HILIC column was used for LC separations. The column temperature was maintained at 40°C. The mobile phase A was Acetonitrile; and B was 10 mM ammonium formate in water at pH 3.2. The gradient was as follows: from 10% to 30% B in 10 min; 100% B from 10 to 12 min; and then back to 10% B at 13 min for 1 min to re-equilibrate prior to the next injection. The flow rate of the mobile phase was 0.3 ml/min, and the injection volume was 5 µl. An electrospray source was employed in positive (for Cer, PE, PC and SM analysis) or negative ion mode (for PI and PS analysis). The collision gas was nitrogen. Needle voltage was set at +4000 V. Several scan modes were used. First, to obtain the naturally different masses of different species, we analysed cell lipid extracts with a precursor ion scan at 184 m/z, 241 m/z, and 264 m/z for PC/SM, PI, and Cer, respectively. We performed a neutral loss scan at 141 and 87 m/z for PE and PS, respectively. The collision energy optimums for Cer, PE, PC, SM, PI, and PS were 25 eV, 20 eV, 30 eV, 25 eV, 45 eV, and 22 eV, respectively. Then, the corresponding SRM transitions were used to quantify different phospholipid species for each class. Two 9 MRM acquisitions were necessary, due to important differences between phospholipid classes. Data were treated with QqQ Quantitative (vB.05.00) and Qualitative analysis software (vB.04.00).

Lipidomic data analysis

Lipidomic analysis was conducted using R sofware computation (R Core Team 2021, URL https://www.R-project.org/). Heatmap was generated with the heatmap.2 function from the R package gplots (version 3.1.1. https://CRAN.R-project.org/package=gplots). Distance matrix computation was done using 1-correlation matrix, method = "complete.obs" Clustering method: WardD2. Principal component analysis was conducted with the PCA function from the R package FactoMineR [15] with default options. Graphical rendering was done using the package factoextra.[16]

^{1}H-NMR based metabolomics

Metabolomic profiling was performed as described previously.[17] All ^{1}H-NMR spectra were obtained on a Bruker DRX-600-Avance NMR spectrometer (Bruker, Wissembourg, France) using the AXIOM metabolomics platform (MetaToul) operating at 600.13 MHz for ^{1}H resonance frequency using an inverse detection 5-mm ^{1}H-^{13}C-^{15}N cryoprobe attached to a cryoplatform (the preamplifier cooling unit).

The ^{1}H-NMR spectra were acquired at 300K using a standard one-dimensional noesypr1D pulse sequence with water presaturation and a total spin-echo delay (2nτ) of 100 ms. A total of 128 transients were collected into 64,000 data points using a spectral width of 12 ppm, a relaxation delay of 2.5 s, and an acquisition time of 2.28 s. ^{1}H-^{1}H COSY, ^{1}H-^{1}H TOCSY, and ^{1}H-^{13}C HSQC were obtained for each biological matrix on one representative sample for metabolite identification.

Data were analysed by applying an exponential window function with a 0.3-Hz line broadening prior to Fourier transformation. The resultant spectra were phased, baseline corrected, and calibrated to TSP (δ 0.00 ppm) manually using Mnova NMR (v9.0, Mestrelab Research). The spectra were subsequently imported into MatLab (R2014a, MathsWorks, Inc.). All data were analysed using full-resolution spectra. The region containing the water resonance (δ 4.6–5.2ppm) was removed, and the spectra were normalized to the probabilistic quotient [18] and aligned using a previously published function.[19]

Data were mean-centred and scaled using the unit variance scaling prior to analysis with orthogonal projection on latent structure-discriminant analysis (O-PLS-DA). ^{1}H-

NMR data were used as independent variables (X matrix) and regressed against a dummy matrix (Y matrix) indicating the class of samples. The O-PLS-derived model was evaluated for goodness of prediction (Q^2Y value) using n-fold cross-validation, where n depends on the sample size. To identify metabolites responsible for discrimination between the groups, the O-PLS-DA correlation coefficients (r^2) were calculated for each variable and back-scaled into a spectral domain so that the shapes of the NMR spectra and the signs of the coefficients were preserved.[20,21] The weights of the variables were colour-coded according to the square of the O-PLS-DA correlation coefficients. Correlation coefficients extracted from significant models were filtered so that only significant correlations above the threshold defined by Pearson's critical correlation coefficient ($P < 0.05$; $|r| > 0.7$; for n=12 per group) were considered significant.

Correlation analysis

Correlation matrixes were created with gene expression levels. Genes were considered as correlated when expression correlation was found to be >0.8 with a p-value <0.05. These correlated genes were assembled in a correlation network. Transcriptional regulatory network were also built using Trrust (https://www.grnpedia.org/trrust/), a manually curated database of mouse transcriptional regulatory network providing information on the interactions between transcription factors (TFs) and their targets and insights into network hubs, motifs and hierarchical organization.[22] Networks of the 50 genes having the highest absolute correlation (Pearson-correlation) with a transcription factor of interest (red node) were displayed using the R function circle Plot.[23] Only the edges corresponding to significant correlations were represented (Bonferroni-adjusted P value < 0.05). Positive and negative correlations were represented by red and blue edges, respectively.

Statistical analysis

Statistical analyses were performed using GraphPad Prism version 7 for Mac OS X (GraphPad Software, San Diego, CA). Two-way ANOVA was performed, followed by appropriate post-hoc tests (Sidak's multiple comparisons test) when differences were found to be significant ($p<0.05$).

Supplementary references

1 Asgharpour A, Cazanave SC, Pacana T, *et al.* A diet-induced animal model of non-alcoholic fatty liver disease and hepatocellular cancer. *J Hepatol* 2016;**65**:579–88. doi:10.1016/j.jhep.2016.05.005

2 Chace DH, Hillman SL, Van Hove JLK, *et al.* Rapid diagnosis of MCAD deficiency: quantitative analysis of octanoylcarnitine and other acylcarnitines in newborn blood spots by tandem mass spectrometry. *Clin Chem* 1997;**43**:2106–13. doi:10.1093/clinchem/43.11.2106

3 Kleiner DE, Brunt EM, Van Natta M, *et al.* Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 2005;**41**:1313–21. doi:10.1002/hep.20701

4 Akpolat N, Yahsi S, Godekmerdan A, *et al.* The value of alpha-SMA in the evaluation of hepatic fibrosis severity in hepatitis B infection and cirrhosis development: A histopathological and immunohistochemical study. *Histopathology* 2005;**47**:276–80. doi:10.1111/j.1365-2559.2005.02226.x

5 Bedossa P, Poitou C, Veyrie N, *et al.* Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. *Hepatology* 2012;**56**:1751–9. doi:10.1002/hep.25889

6 Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 1987;**162**:156–9. doi:10.1016/0003-2697(87)90021-2

7 Smyth GK. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Stat Appl Genet Mol Biol* 2004;**3**:1–25. doi:10.2202/1544-6115.1027

8 Klipper-Aurbach Y, Wasserman M, Braunspiegel-Weintrob N, *et al.* Mathematical formulae for the prediction of the residual beta cell function during the first two years of disease in children and adolescents with insulin-dependent diabetes mellitus. *Med Hypotheses* 1995;**45**:486–90.http://www.ncbi.nlm.nih.gov/pubmed/8748093 (accessed 19 Jul 2019).

9 Szklarczyk D, Gable AL, Lyon D, *et al.* STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019;**47**:D607–13. doi:10.1093/nar/gky1131

10 Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 2002;**30**:207–10. doi:10.1093/nar/30.1.207

11 Bolstad BM, Irizarry RA, Astrand M, *et al.* A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003;**19**:185–93.

12 Klipper-Aurbach Y, Wasserman M, Braunspiegel-Weintrob N, *et al.* Mathematical formulae for the prediction of the residual beta cell function during the first two years of disease in children and adolescents with insulin-

dependent diabetes mellitus. *Med Hypotheses* 1995;**45**:486–90.

13 Mardinoglu A, Shoaie S, Bergentall M, *et al.* The gut microbiota modulates host amino acid and glutathione metabolism in mice. *Mol Syst Biol* 2015;**11**:834–834. doi:10.15252/msb.20156487

14 Bligh EG, Dyer WJ. A rapid method of total lipid extraction and purification. *Can J Biochem Physiol* 1959;**37**:911–7. doi:10.1139/y59-099

15 Lê S, Josse J, Husson F. FactoMineR: An R package for multivariate analysis. *J Stat Softw* 2008;**25**:1–18. doi:10.18637/jss.v025.i01

16 Kassambara A. 'ggplot2' Based Publication Ready Plots [R package ggpubr version 0.4.0]. 2020.

17 Beckonert O, Keun HC, Ebbels TMD, *et al.* Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc* 2007;**2**:2692–703. doi:10.1038/nprot.2007.376

18 Dieterle F, Ross A, Schlotterbeck G, *et al.* Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal Chem* 2006;**78**:4281–90. doi:10.1021/ac051632c

19 Veselkov KA, Lindon JC, Ebbels TMD, *et al.* Recursive Segment-Wise Peak Alignment of Biological ^{1}H NMR Spectra for Improved Metabolic Biomarker Recovery. *Anal Chem* 2009;**81**:56–66. doi:10.1021/ac8011544

20 Cloarec O, Dumas ME, Trygg J, *et al.* Evaluation of the Orthogonal Projection on Latent Structure Model Limitations Caused by Chemical Shift Variability and Improved Visualization of Biomarker Changes in ^{1}H NMR Spectroscopic Metabonomic Studies. *Anal Chem* 2005;**77**:517–26. doi:10.1021/ac048803i

21 Cloarec O, Dumas M-E, Craig A, *et al.* Statistical Total Correlation Spectroscopy: An Exploratory Approach for Latent Biomarker Identification from Metabolic ^{1}H NMR Data Sets. *Anal Chem* 2005;**77**:1282–9. doi:10.1021/ac048630x

22 Han H, Cho J-W, Lee S, *et al.* TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res* 2018;**46**:D380–6. doi:10.1093/nar/gkx1013

23 Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 2008;**9**:559. doi:10.1186/1471-2105-9-559