Science Translational Medicine
AAAS

Supplementary Materials for

Risk assessment with gut microbiome and metabolite markers in NAFLD development

Howell Leung *et al.*

Corresponding author: Yueqiong Ni, yueqiong.ni@leibniz-hki.de; Huating Li, huarting99@sjtu.edu.cn; Weiping Jia, wpjia@sjtu.edu.cn; Gianni Panagiotou, gianni.panagiotou@hki-jena.de

Sci. Transl. Med. **14**, eabk0855 (2022)
DOI: 10.1126/scitranslmed.abk0855

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

MATERIALS AND METHODS

Anthropometric and biochemical measurements

All participants underwent basic physical examinations with routine biochemical blood analysis, as described previously (*94*, *95*). Anthropometric parameters including body weight, height, body mass index (BMI), systolic blood pressure (SBP) and diastolic blood pressure (DBP) were measured by trained investigators at local community clinics. Fasting blood and urine samples were collected for measurement of biochemical indexes. Plasma concentrations of fibroblast growth factor 21 (FGF21) were measured using ELISA kits (Antibody and Immunoassay Services, The University of Hong Kong). Haemoglobin A1c (HbA1c) levels in red blood cells were measured by high-performance liquid chromatography using a Bio-Rad Variant II System (Bio-Rad Laboratories). Participants underwent a 75g oral glucose tolerance test for glucose and insulin measurements. Basal insulin secretion and insulin sensitivity were estimated by the homeostasis model assessment (HOMA): (HOMA-β = FINS [mU/L] × 20)/(FPG [mmol/L] -3.5) and HOMA-IR = FINS (mU/L) × FPG (mmol/L)/22.5, where FINS is fasting insulin, FPG is fasting plasma glucose, and IR is insulin resistance. Demographics, smoking habits, alcohol consumption, family history of diseases, history of surgery, and current and past medical therapy were obtained through a standard questionnaire. Abdominal ultrasonography was performed by experienced sonographers who were blinded to clinical presentation and laboratory findings.

Clinical diagnosis of NAFLD

Clinical diagnosis of NAFLD followed the criteria proposed by the Asian Pacific Association for the Study of the Liver (APASL) (*37*), one of the leading associations for investigation and

treatment of liver diseases in the world, based on B ultrasonography showing a diffuse increase in fine echoes in the liver parenchyma compared with the kidney or spleen parenchyma, ruling out secondary causes of hepatic fat accumulation including acute infectious disease, biliary obstructive diseases, alcohol abuse (more than 140g of ethanol/week for men or 70g of ethanol/week for women), acute or chronic cholecystitis, acute or chronic viral hepatitis, cirrhosis.

Targeted metabolomics profiling

Metabolomics analysis targeting gut microbiota-host co-metabolism (*96*) was performed by Metabo-Profile. Samples were treated according to the following steps: cell samples harvested and stored in Eppendorf Safelock microcentrifuge tubes were mixed with 10 prechilled zirconium oxide beads and 20 μL of deionized water. Samples were homogenized for 3 min and 150 μL of methanol containing internal standards was added to extract metabolites. Samples were homogenized for another 3 min and then centrifuged at 18,000 × g for 20 min. Supernatants were transferred to a 96-well plate. The following procedures were performed on a Biomek 4000 workstation (Biomek 4000, Beckman Coulter, Inc.). 20 μL of freshly prepared derivatization reagents were added to each well. Plates were sealed and the derivatization was at 30°C for 60 min. After derivatization, samples were evaporated for 2 h and 330 μL ice-cold 50% methanol solution was added to reconstitute samples. Plates were stored at -20°C for 20 min followed by 4,000 × g centrifugation at 4°C for 30 min. 135 μL of supernatant was transferred to a new 96-well plate with 10 μL internal standards in each well. Serial dilutions of derivatized stock standards were added to the wells left. Last, the plate was sealed for LC-MS analysis. All standards were from Sigma-Aldrich, Steraloids Inc. and TRC Chemicals. All standards were accurately weighed and prepared in water, methanol, sodium hydroxide solution, or hydrochloric acid solution to obtain

individual stock solutions at 5.0 mg/mL. Appropriate amounts of stock solutions were mixed to create stock calibration solutions.

An ultraperformance liquid chromatography coupled to tandem mass-spectrometry (UPLC-MS/MS) system (ACQUITY UPLC-Xevo TQ-S, Waters Corp.) was used to quantitate all targeted metabolites by Metabo-Profile Biotechnology Co., Ltd. The optimized instrument settings are briefly described as follows. For HPLC, column: ACQUITY HPLC BEH C18 1.7×10^{-6} m VanGuard pre column (2.1 × 5 mm) and ACQUITY HPLC BEH C18 1.7×10^{-6} m analytical column (2.1 × 100 mm), column temp.: 40°C, sample manager temp.: 10°C, mobile phases: A = water with 0.1% formic acid and B = acetonitrile/IPA (70:30), gradient conditions: 0–1 min (5% B), 1–11 min (5–78% B), 11–13.5 min (78–95% B), 13.5–14 min (95–100% B), 14–16 min (100% B), 16–16.1 min (100-5% B), 16.1–18 min (5% B), flow rate: 0.40 mL/min and injection vol.: 5.0 μL. For mass spectrometer, capillary: 1.5 (ESI+), 2.0 (ESI-) Kv, source temp.: 150°C, desolvation temp.: 550°C, and desolvation gas flow: 1,000 L/h. The raw data generated by UPLC-MS/MS were processed using iMAP platform (v1.0; Metabo-Profile) to calculate concentrations of analytes in the samples.

Stool sample collection and DNA extraction

Participants were given a commercial tube containing DNA stabilizer (STRATEC Molecular) to collect stool samples at home. After being returned to the study staff, samples were stored at –80 °C until extraction. Stool DNA was extracted using PSP Spin Stool DNA Kits (STRATEC Molecular) according to the manufacturer's instructions.

Library preparation and shotgun metagenomic sequencing

Sequencing library was generated based on Illumina technologies and following manufacturers' recommendations. Index codes were added to each sample. Genomic DNA was randomly fragmented to 350 bp and DNA fragments were narrowly size selected with sample purification beads. The selected fragments were then end-polished, A-tailed, and ligated with adapter. These fragments were filtered with beads again and amplified by PCR. At last, the library was analysed for size distribution and quantified using real-time PCR. The library was then sequenced on an Illumina platform Novaseq 6000 (Novogene) with paired-end reads of 150 bp.

Metagenomics profiling

Quality control and removal of host-derived reads

Quality control was as described previously (*97*). All Illumina primer/adapter/linker sequences were removed and low-quality regions (consecutive regions with Phred quality <20) were subsequently trimmed. All remaining reads were mapped to the human genome with BWA version 0.7.4 (*98*) and reads with >95% identity and 90% coverage were removed as human DNA contamination.

Taxonomy profiling and functional annotation

Taxonomic annotation of the high-quality reads was by MetaPhlAn2 (*42*) version 2.7.7 with default settings, generating taxonomic relative abundances. Bacterial community profiles were constructed at the family, genus and species level for further analyses. Functional annotation of the high-quality reads after quality control was by HUMAnN2 (*48*) pipeline version 0.11.2. The quantified pathway and gene family abundances in units of reads per kilobase were normalized to copies per million units, resulting in transcript-per-million-like normalization. MGS (*43*) pipeline with

standard settings was applied to cluster gene family abundances into co-abundant gene groups, as an alternative taxonomic profile.

Diversity analysis for metagenomics and metabolomics data

Alpha diversity indices Shannon, Simpson and Chao1 were calculated using the R packages vegan (*99*) and fossil (*100*) based on relative abundance at the family, genus and species level. For estimating community dissimilarities, Bray-Curtis distances were calculated using the R package vegan.

Metagenomics and metabolomics data preprocessing

Prevalence filtering was carried out to remove low-prevalence features, prevalence <10%, before differentially abundant analysis and diversity analysis. Additional cumulative sum scaling (CSS) normalization using R package metagenomeSeq (*101*) was applied only to taxonomic data and functional data.

Data preprocessing prior to machine learning

Prevalence filtering was used to remove low-prevalence features (prevalence <10%). In addition, CSS normalization using the R package metagenomeSeq (*101*) was applied only to taxonomic data and functional data. We removed zero- and near zero-variance features, and scaled the data to the interval between 0 and 1 prior to feature selection and model training, using R package caret (*102*).

Feature selection

Feature selection was done separately on taxonomic data, functional data, metabolomic data and clinical metadata, as they had different characteristics. Univariate feature selection was chosen, as it gave the best performance in accuracy and stability among the feature selection methods we examined. Different univariate feature selection methods (zero-inflated Gaussian mixture model for taxonomic data and functional data; generalized linear model with inverse gamma distribution for metabolomic data and Wilcoxon rank-sum test for clinical metadata) and P-value thresholds were applied independently on the four types of data to identify features for discriminating $NAFLD^{-/-}$ and $NAFLD^{-/+}$ groups. To further improve our noninvasive microbiome model, we examined the minimum number of additional noninvasive clinical features needed for a significant improvement, and manually added anthropometric parameters BMI and age to further improve the microbiome model.

Model building and model evaluation

A random forest classifier was built to classify participants into $NAFLD^{-/-}$ and $NAFLD^{-/+}$ based on the combination of metagenomic, metabolomic and clinical data. Model implementation and performance evaluation were done using the R packages caret, ranger and MLeval (*102-104*).

A leave-one-out iterative approach was applied to build the final model, which included the following steps: (1) Combine taxonomic data (77 genera), functional data (382 pathways), metabolomic data (123 metabolites) and non-microbiome data (2 anthropometrics parameters and 25 noninvasive clinical parameters) into one dataset (n = 180, number of features = 609); (2) remove one participant and use the rest as the training data (n = 179); (3) apply a prevalence filter and perform feature selection on the training dataset to select the important features; (4) train a random forest model using training data and the selected features with 5 repeated 5-fold cross

validation for hyperparameter tuning; (5) test the training model with best tunings on the one participant removed in step 2; (6) repeat steps 2-5 180 times and summarize the predicted scores of all samples to obtain the overall testing performances, and calculate the receiver operating characteristic curve (ROC), precision-recall curve, and confusion matrix (generated with the optimal probability cutoff of the ROC curve); and (7) apply steps 3-4 to the entire dataset (n = 180) to obtain the final model, with performance estimated as in step 6.

We applied the above leave-one-out iterative approach to build 3 models: the first model had only metagenomics and metabolomics data; the second added 2 noninvasive clinical parameters; and the third added 2 anthropometric parameters and 2 noninvasive clinical parameters. Differences of areas under the ROC (auROCs) was assessed using the roc.test function from R package pROC (*105*).

Prognosis performance of previous clinical models

Previous clinical prospective NAFLD studies (*19-21*) proposed three clinical models for predicting NAFLD development over the long term, including fibroblast growth factor 21 and body mass index (FGF21 + BMI), fatty liver index (FLI), and triglyceride and glucose index (TyG). To compare the prognostic performance of our model with these clinical models, we built 3 models with our data, using the clinical parameters only. For each model, we applied the leave-one-out iterative approach using logistic regression. To assess the importance of the microbiome and metabolome, we also used a leave-one-out iterative approach (with random forest instead of logistic regression) to build 3 models combining our selected microbiome features with these clinical parameters. DeLong test was performed using the roc.test function from R package pROC (*105*) to compare the auROCs between models.

Evaluating the diagnostic ability of the model in external cohorts

To our knowledge, no similar studies have conducted long-term follow up of NAFLD development in healthy individuals using a combination of gut metagenome, metabolome and clinical features as a risk-assessment tool. Thus, we were unable to test our prospective model directly in an external cohort. Instead, we used external case-control cohorts to examine the ability of our final prognostic model to classify correctly NAFLD and healthy participants. Four cohorts were used, including two cohorts of Chinese: (1) 78 patients with NAFLD and 10 controls without NAFLD, as diagnosed with biopsy (BioProject ID: PRJNA732131); and (2) 111 MRS-diagnosed NAFLD patients and 8 controls (BioProject IDs: PRJNA703757 and PRJNA414688); and two biopsy-diagnosed cohorts of other ethnicity: (3) a European cohort of 46 patients with NAFLD and 10 controls (*54*), and (4) a US cohort of 26 cirrhosis patients and 54 controls (*49*). Further additional data (e.g. anthropometric and/or available clinical data) for the two Chinese validation cohorts besides grouping information, please contact the corresponding author.

Because some selected features included in the final model were not available in the external cohorts, we were unable to test our model directly. Instead, we built a new prognostic model based on the $NAFLD^{-/+}$ and $NAFLD^{-/-}$ groups using a subset of the 18 selected features that were available in the external cohorts. In the model for the two Chinese cohorts and the European cohort, 9 of the 18 selected features were used: 2 genera, 3 pathways, 2 anthropometric parameters and 2 non-invasive clinical metadata; while 7 of the 18 selected features were used in the model for the US cohort: 2 genera, 1 pathway, 2 anthropometric parameters and 2 noninvasive clinical metadata. Performances of models, including ROC curves, precision-recall curves and confusion

matrices (generated with the optimal probability cutoff of the ROC curve), were produced by applying the model to the unseen external cohort data.

SHAP-based model interpretation

SHapley Additive exPlanations (SHAP), based on Shapley values, is a concept from game theory for revealing feature attributions in any model (*65, 106*). We adapted the method described in Artzi et al. (*107*) to explain our prospective model. The SHAP values of our prospective model were computed using R package fastshap (*108*). The SHAP values of each feature represented their contribution to the model prediction. Positive SHAP values indicated prediction of $NAFLD^{-/+}$ and negative SHAP values indicated prediction of $NAFLD^{-/-}$. Feature importance was calculated by averaging the absolute SHAP values of each feature. The contribution of each feature set was also computed by summing the SHAP values for each feature category, then taking the mean of the absolute sum across all samples. The optimal threshold of each feature was defined according to the inflection point in the SHAP dependence plot. We also transformed each feature into binary variables (above and below threshold), and tested their associations with the groupings ($NAFLD^{-/-}$ and $NAFLD^{-/+}$) using chi-square tests.

Classifying different degrees of steatosis and change of fibrosis

To distinguish between severe and mild steatosis, we computed liver fat percentage (*109*) as an estimation of liver fat content. Because there is no cutoff for liver fat percentage to define mild or severe steatosis, we used upper and lower tertiles of liver fat percentage for $NAFLD^{-/+}$ participants at the time of diagnosis (4.6 years after enrollment) to obtain the mild and severe groups. Another random forest model to classify mild and severe cases was then built using the features from our

final prospective model (fasting insulin was not included because it is used in the calculation of liver fat percentage).

To classify different degrees of fibrosis change, we first computed the fibrosis 4 (FIB-4) scores (*110*) as laboratory values of fibrosis for both 2014 and 2018, as well as the change of FIB-4 scores after 4 years. Then two groups namely "FIB-4-up" and "FIB-4-down" were selected from the $NAFLD^{-/+}$ group, based on the 1^{st} quintile and 5^{th} quintile of FIB-4 changes from baseline to follow-up. A new model based on baseline features to classify different degrees of fibrosis change was then trained (after feature selection) and evaluated using similar approach as the prospective model described above.

Data visualization

All figures were generated by R software 3.6.3, using ggplot2 and ComplexHeatmap packages.

SUPPLEMENTARY FIGURES

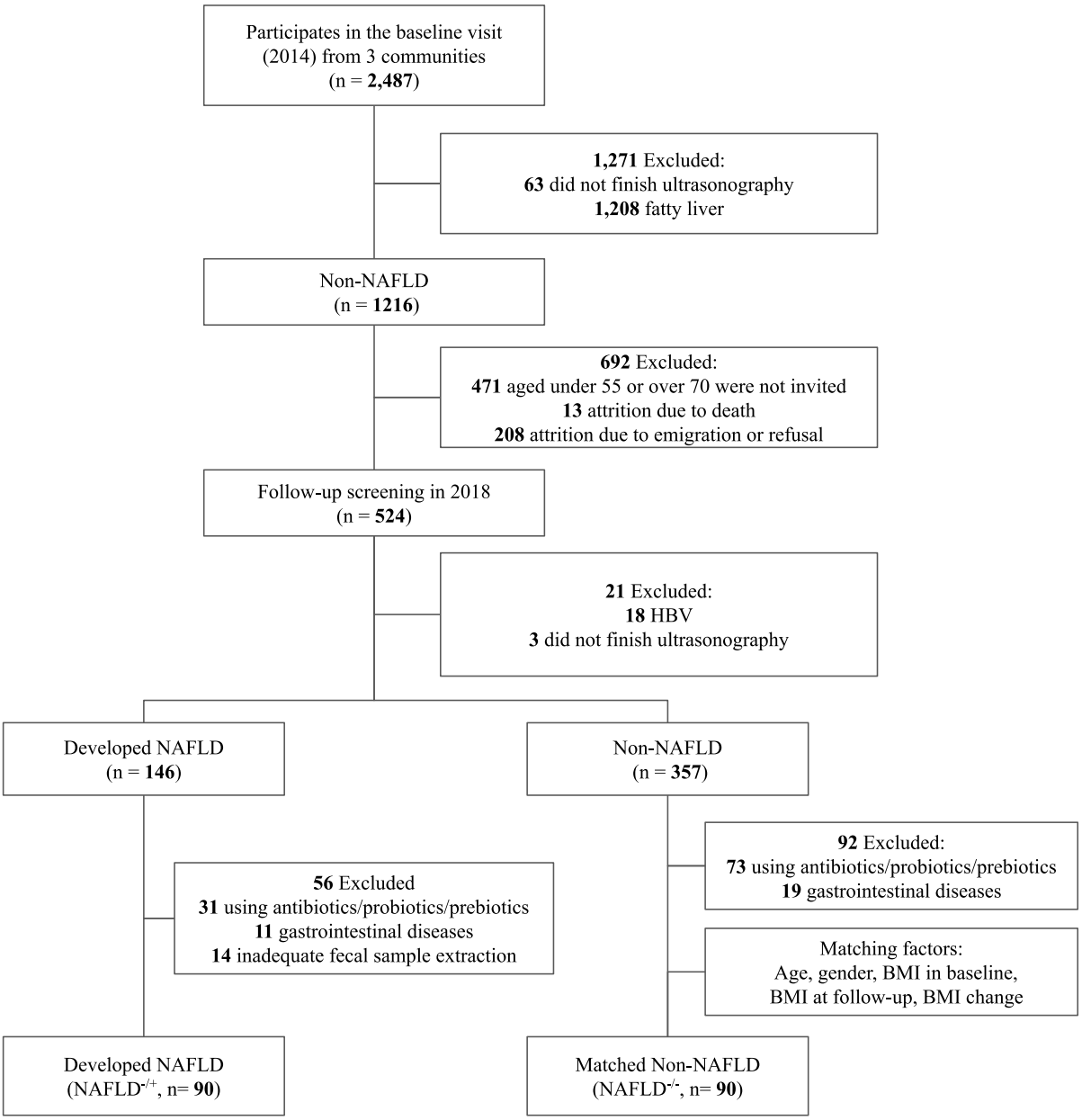

Figure S1. Flow chart of participant enrolment.

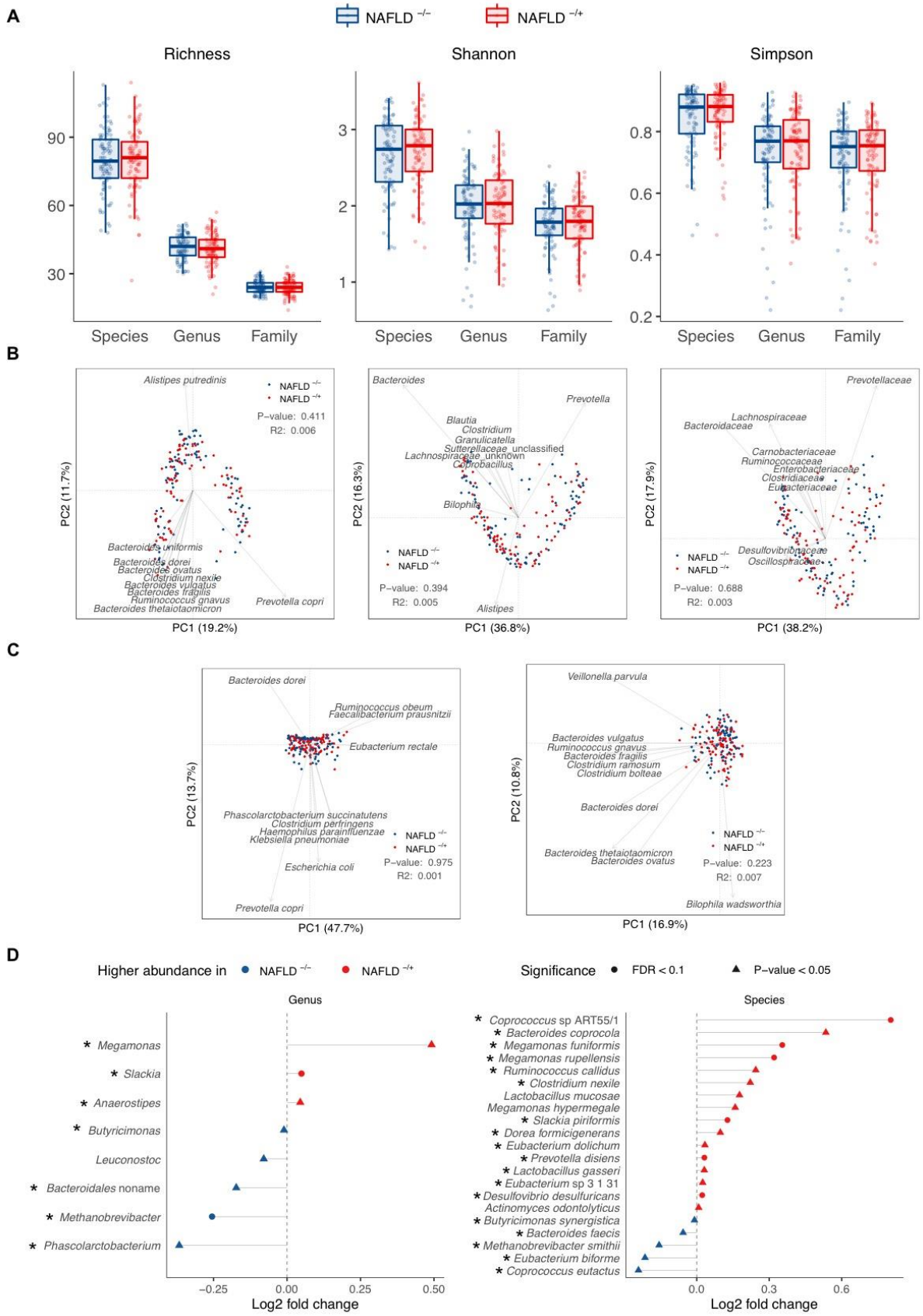
A
NAFLD −/−
NAFLD −/+
Richness
Shannon
Simpson
Species
Genus
Family
B
Alistipes putredinis
Bacteroides uniformis
Bacteroides dorei
Bacteroides ovatus
Clostridium nexile
Bacteroides vulgatus
Bacteroides fragilis
Ruminococcus gnavus
Bacteroides thetaiotaomicron
Prevotella copri
P−value: 0.411
R2: 0.006
PC1 (19.2%)
PC2 (11.7%)
Bacteroides
Prevotella
Blautia
Clostridium
Granulicatella
Sutterellaceae_unclassified
Lachnospiraceae_unknown
Coprobacillus
Bilophila
Alistipes
P−value: 0.394
R2: 0.005
PC1 (36.8%)
PC2 (16.3%)
Prevotellaceae
Lachnospiraceae
Bacteroidaceae
Carnobacteriaceae
Ruminococcaceae
Enterobacteriaceae
Clostridiaceae
Eubacteriaceae
Desulfovibrionaceae
Oscillospiraceae
P−value: 0.688
R2: 0.003
PC1 (38.2%)
PC2 (17.9%)
C
Bacteroides dorei
Ruminococcus obeum
Faecalibacterium prausnitzii
Eubacterium rectale
Phascolarctobacterium succinatutens
Clostridium perfringens
Haemophilus parainfluenzae
Klebsiella pneumoniae
Escherichia coli
Prevotella copri
P−value: 0.975
R2: 0.001
PC1 (47.7%)
PC2 (13.7%)
Veillonella parvula
Bacteroides vulgatus
Ruminococcus gnavus
Bacteroides fragilis
Clostridium ramosum
Clostridium bolteae
Bacteroides dorei
Bacteroides thetaiotaomicron
Bacteroides ovatus
Bilophila wadsworthia
P−value: 0.223
R2: 0.007
PC1 (16.9%)
PC2 (10.8%)
D
Higher abundance in
NAFLD −/−
NAFLD −/+
Significance
FDR < 0.1
P−value < 0.05
Genus
* Megamonas
* Slackia
* Anaerostipes
* Butyricimonas
Leuconostoc
* Bacteroidales noname
* Methanobrevibacter
* Phascolarctobacterium
−0.25
0.00
0.25
0.50
Log2 fold change
Species
* Coprococcus sp ART55/1
* Bacteroides coprocola
* Megamonas funiformis
* Megamonas rupellensis
* Ruminococcus callidus
* Clostridium nexile
Lactobacillus mucosae
Megamonas hypermegale
* Slackia piriformis
* Dorea formicigenerans
* Eubacterium dolichum
* Prevotella disiens
* Lactobacillus gasseri
* Eubacterium sp 3 1 31
* Desulfovibrio desulfuricans
Actinomyces odontolyticus
* Butyricimonas synergistica
* Bacteroides faecis
* Methanobrevibacter smithii
* Eubacterium biforme
* Coprococcus eutactus
0.0
0.3
0.6
Log2 fold change

Figure S2. Comparison of the gut microbiome taxonomy profiles of NAFLD$^{-/-}$ and NAFLD$^{-/+}$ using MetaPhlAn2. (**A**) Alpha diversities of gut microbiome species, genera and families with Chao1 index (left), Shannon index (middle) and Simpson index (right). Box plots showed median (center lines), lower/upper quartiles (box limits) and whiskers (the last data points 1.5 times interquartile range (IQR) from the lower or upper quartiles). (**B**) Principal coordinate analysis (PCoA) of Bray-Curtis dissimilarity between gut microbiome abundance profiles, at species, genus and family level. (**C**) PCoA of weighted (left) and unweighted (right) UniFrac distance between gut microbiome abundance profiles at the species level. (**D**) Log2-fold change of significantly differentially abundant genus and species (P-value < 0.05, zero-inflated Gaussian mixture model). Genera or species that remained significant after adjustment for age, gender, BMI, and HOMA-IR are indicated by asterisks. NAFLD, nonalcoholic fatty liver disease; BMI, body mass index; HOMA-IR, homeostasis model assessment for insulin resistance; FDR, false discovery rate.

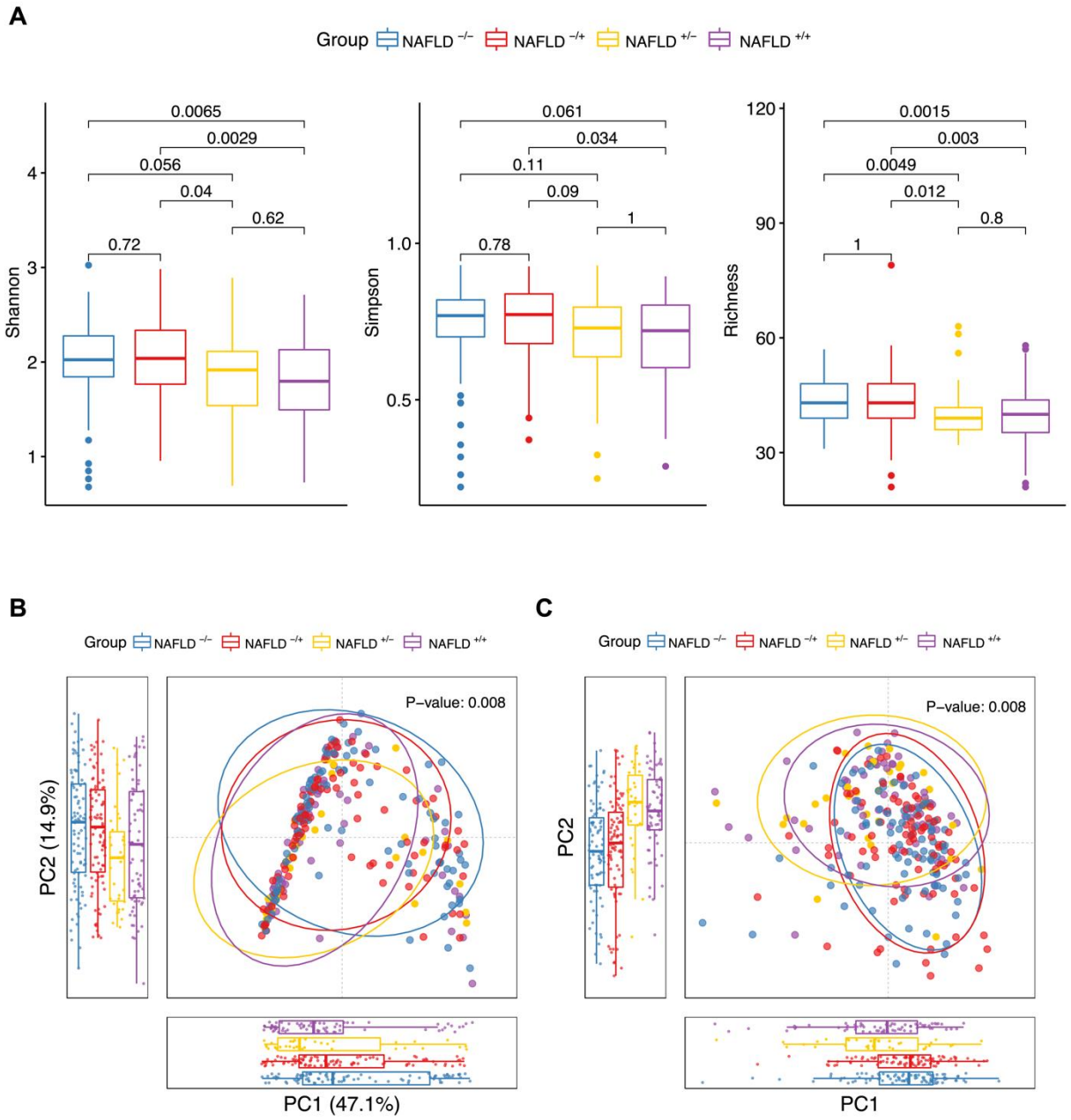

Figure S3. Comparison of the gut microbiome taxonomy profiles among four groups using MetaPhlAn2. (**A**) Alpha diversities of gut microbiome genera with Shannon index (Shannon) and Simpson index (middle) and Chao1 index (right). Box plots showed median (center lines), lower/upper quartiles (box limits) and whiskers (the last data points 1.5 times interquartile range (IQR) from the lower or upper quartiles). (**B**) Principal coordinate analysis (PCoA) and (**C**)

principal component analysis (PCA) of Bray-Curtis dissimilarity between gut microbiome abundance profiles at genus level.

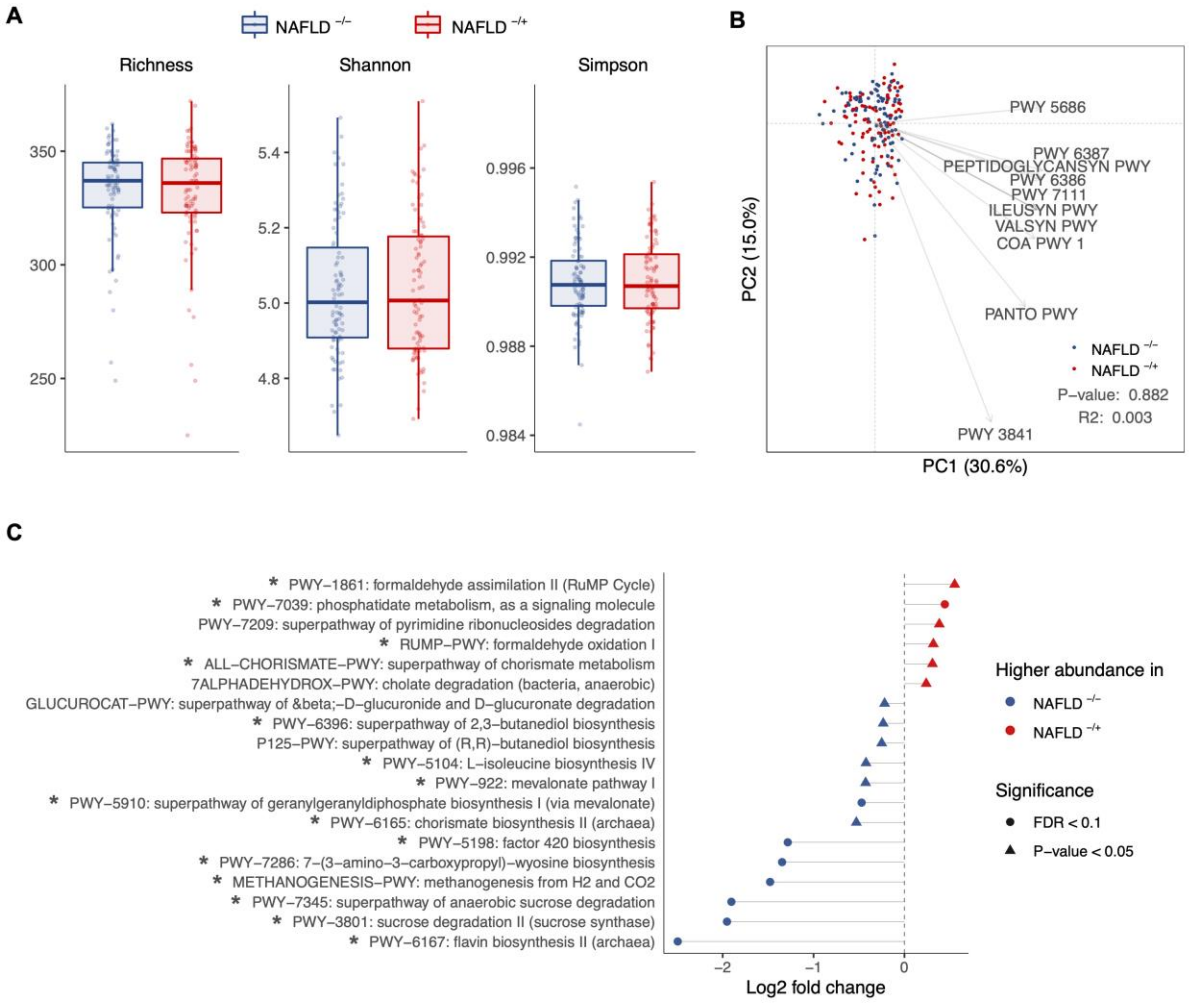

Figure S4. Comparison of gut microbiome functional profiles of NAFLD$^{-/-}$ and NAFLD$^{-/+}$. (**A**) Alpha diversities of gut microbiome at the MetaCyc pathway level by Chao1 index (left), Shannon index (middle) and Simpson index (right). Box plots showed median (center lines), lower/upper quartiles (box limits) and whiskers (the last data points 1.5 times interquartile range (IQR) from the lower or upper quartiles). (**B**) Principal coordinate analysis of Bray-Curtis dissimilarity between gut microbiome MetaCyc pathways profiles. (**C**) Log2-fold change of significantly differentially abundant pathways (P-value < 0.05, zero-inflated Gaussian mixture model). Pathways that remained significant after adjustment for age, gender, BMI, and HOMA-IR are

indicated by asterisks. NAFLD, nonalcoholic fatty liver disease; BMI, body mass index; HOMA-IR, homeostasis model assessment for insulin resistance; FDR, false discovery rate.

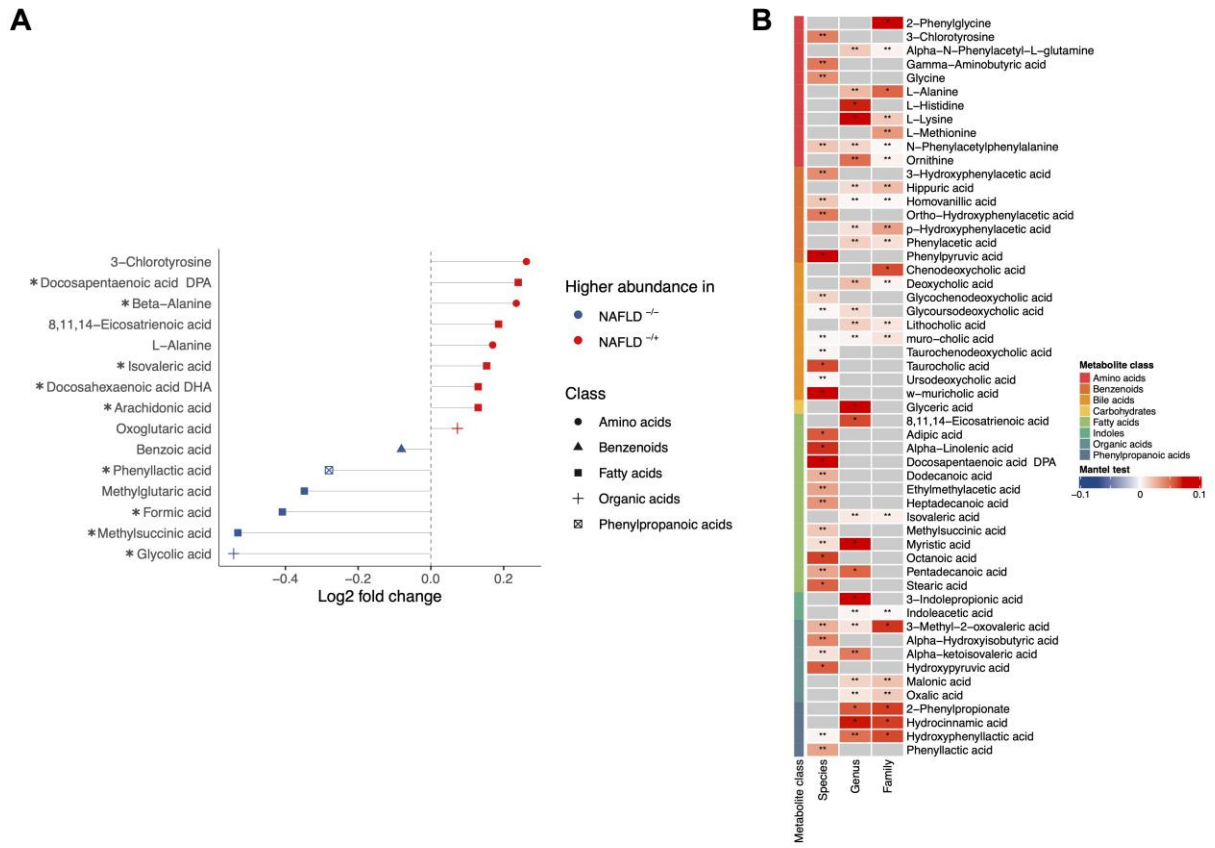

Figure S5. Comparison of serum metabolome profiles of NAFLD$^{-/-}$ and NAFLD$^{-/+}$, and correlations between metabolites and gut microbiota taxonomic profiles. (**A**) Log2-fold change (using median) of significantly differentially abundant metabolites (P-value < 0.05, generalized linear model). Classes of metabolites are indicated by shape. Metabolites that remained significant after adjustment for age, gender, BMI, and HOMA-IR are indicated by asterisks. NAFLD, nonalcoholic fatty liver disease; BMI, body mass index; HOMA-IR, homeostasis model assessment for insulin resistance. (**B**) Mantel tests (based on Spearman's correlation coefficients) were used to analyse the correlations of metabolites with the taxonomic profiles at species, genus, and family level. Colours of cells indicate the direction and magnitude of significant correlations, with grey representing non-significant correlations. *, P-value < 0.1; **, P-value < 0.05.

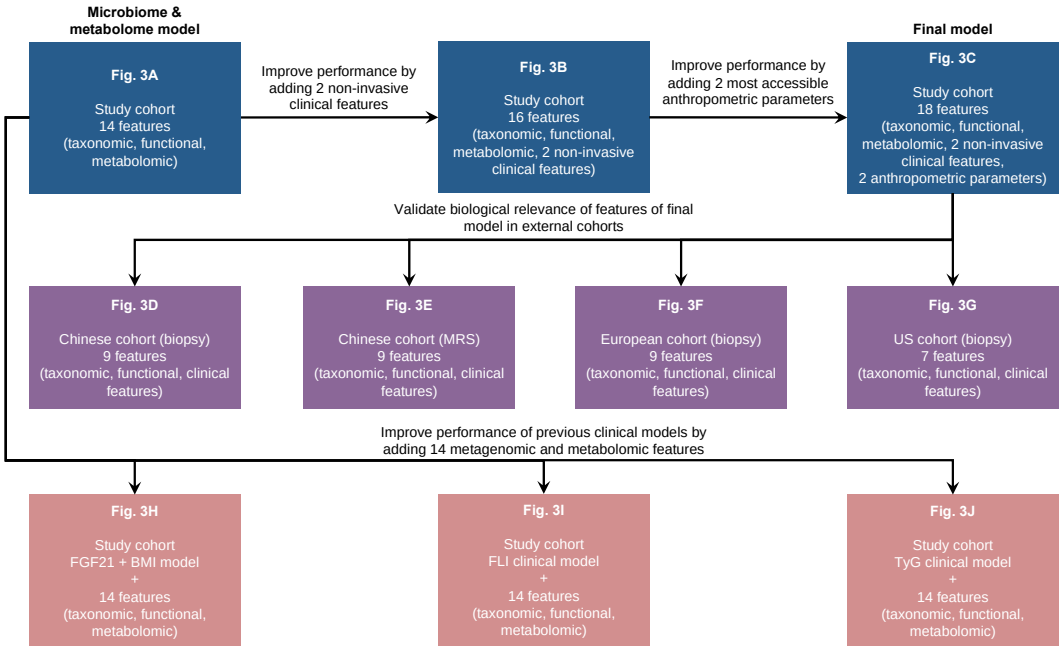

Figure S6. Flowchart showing the overall machine learning analysis framework. The cohorts and data features that were used at different stages of the analysis are shown. Blue boxes are models discriminating between NAFLD-/+ and NAFLD-/- in our study cohort; purple boxes are validation models built on our study cohort and tested in external cohorts discriminating healthy and NAFLD; peach boxes are models built on this study cohort, comparing performance of previous clinical models with and without our metagenomics + metabolome features.

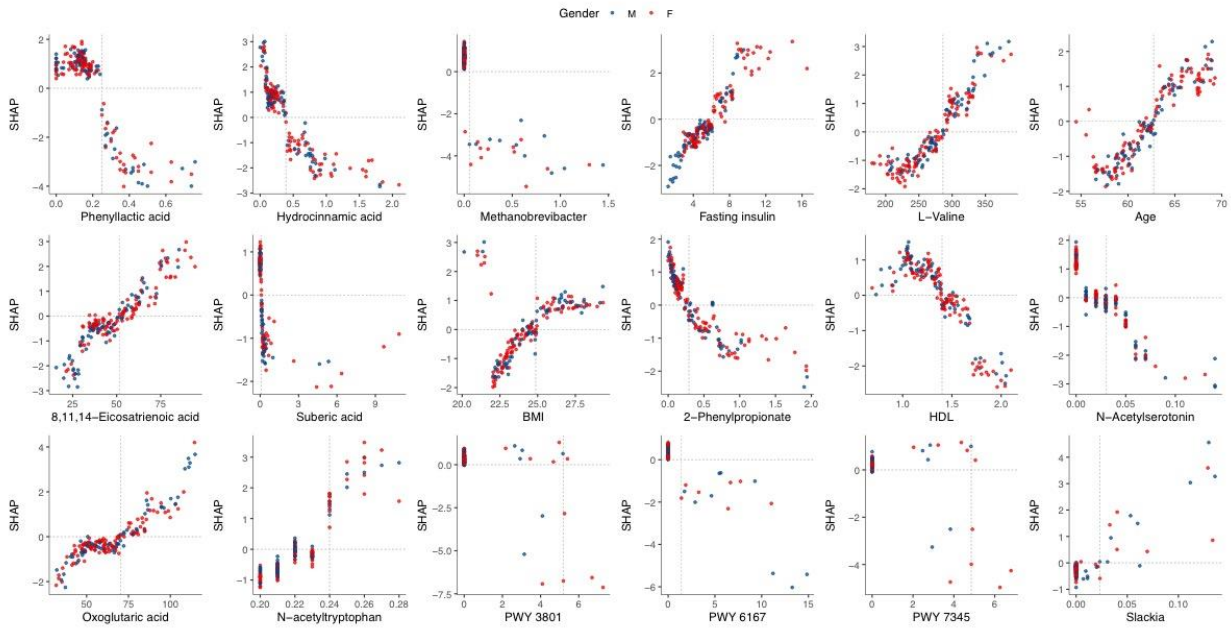

Figure S7. Dependence plots of all 18 selected features in the final model. Dependence plots showing the effect of each feature on model prediction. Extreme values were dropped for better visualisation. Plots are in descending order by their contribution in the model, from left to right and top to bottom. Each point represents a participant (n = 180). Colour indicates sex with blue for male and red for female. The x-axis is the feature value and the y-axis is the SHAP value for the feature. Negative SHAP value indicates the feature attribution for prediction of NAFLD$^{-/-}$ and positive SHAP value indicates the feature attribution for prediction of NAFLD$^{-/+}$. Optimal thresholds for features are indicated by vertical dotted lines. NAFLD, nonalcoholic fatty liver disease; SHAP, SHapley Additive exPlanations.

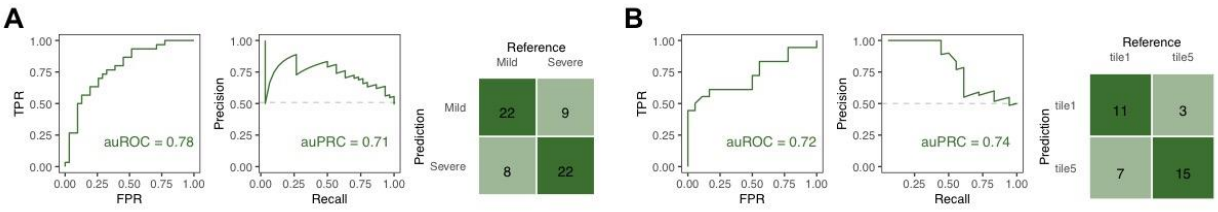

Figure S8. Performance of ML models classifying different degrees of steatosis and change of fibrosis.

Performance of leave-one-out iterative machine learning models: (**A**) a model discriminating mild and severe steatosis cases using metagenomics + metabolomic features from our prospective model, and (**B**) a model classifying subjects by the fibrosis deterioration using metagenomics features.

Table S1. Comorbidities and use of medications in participants at baseline or follow-up visit.

	Baseline			Follow-up		
	NAFLD-/- (n=90)	**NAFLD-/+ (n=90)**	**P value**	**NAFLD-/- (n=90)**	**NAFLD-/+ (n=90)**	**P value**
Diabetes	13 (14.4%)	19 (21.1%)	0.2421	14 (15.6%)	19 (21.1%)	0.3355
Hypertention	30 (33.3%)	40 (44.4%)	0.1263	49 (54.4%)	55 (61.1%)	0.3652
Metabolic syndrome	13 (14.4%)	23 (25.6%)	0.0624	40 (44.4%)	56 (62.2%)	0.0168
Antidiabetic drugs	5 (5.5%)	6 (6.7%)	0.7557	7 (7.8%)	8 (8.9%)	0.7874
Antihyperlipidemic drugs	0 (0.0%)	0 (0.0%)	-	1 (1.1%)	5 (5.6%)	0.0911

*P-value denotes differences between NAFLD-/+ and NAFLD-/- analyzed by Chi-square test.

Table S2. Comparison of gut microbiome taxonomy profiles of NAFLD$^{-/-}$ and NAFLD$^{-/+}$ using MGS. Alpha and beta diversities of the gut microbiome at species, genus, and family level were compared. Wilcoxon rank-sum test was used to compare alpha diversities; permutational multivariate analysis of variance was used to analyse beta diversities. P values are shown.

	Alpha diversity			Beta diversity
	Richness	**Shannon**	**Simpson**	**Bray-Curtis Distance**
species	0.47	0.076	0.087	0.196
genus	0.8	0.3	0.31	0.103
family	0.92	0.59	0.48	0.09

Table S3. Zero-inflated gaussian mixture model results for enzymes (EC), with and without adjustment of age, gender, BMI and HOMA-IR. Positive log2 fold change indicates NAFLD$^{-/+}$ > NAFLD$^{-/-}$, negative indicates NAFLD$^{-/-}$ > NAFLD$^{-/+}$. P-values and FDR adjusted P-values are shown. Only significant ECs based on p-value without adjustment of age, gender, BMI, and HOMA-IR are shown here.

EC	log2FC	*P* value (no adj.)	FDR (no adj.)	*P* value (adj. with age, gender, BMI and HOMA-IR)	FDR (adj. with age, gender, BMI and HOMA-IR)
EC_2.1.1.86	-2.02	5.98E-14	9.65E-11	1.63E-11	2.63E-08
EC_1.12.98.1	-1.53	5.68E-12	4.58E-09	1.42E-09	7.65E-07
EC_2.8.4.1	-0.862	2.77E-11	1.49E-08	3.35E-10	2.71E-07
EC_2.7.1.161	-0.696	9.58E-11	3.85E-08	1.03E-08	3.32E-06
EC_5.4.99.13	-1.25	1.19E-10	3.85E-08	7.97E-09	3.22E-06
EC_2.3.1.101	-0.675	1.46E-10	3.92E-08	3.46E-08	6.99E-06
EC_2.5.1.46	-0.574	1.94E-10	4.48E-08	1.56E-07	2.79E-05
EC_3.5.4.27	-0.573	3.71E-10	6.24E-08	7.07E-06	4.96E-04
EC_3.5.1.102	-0.595	3.85E-10	6.24E-08	1.96E-06	2.10E-04
EC_2.5.1.77	-0.867	4.67E-10	6.24E-08	2.08E-05	1.16E-03
EC_3.5.4.41	-0.654	4.85E-10	6.24E-08	2.65E-06	2.38E-04
EC_1.12.98.2	-0.616	4.93E-10	6.24E-08	2.21E-06	2.10E-04
EC_2.1.1.216	-0.611	5.02E-10	6.24E-08	2.37E-07	3.48E-05
EC_3.1.27.9	-0.585	5.56E-10	6.41E-08	1.53E-08	4.12E-06
EC_1.5.98.2	-0.594	6.24E-10	6.71E-08	9.65E-07	1.30E-04
EC_1.1.1.34	-0.61	7.13E-10	6.81E-08	6.83E-06	4.96E-04
EC_1.5.98.1	-0.662	7.17E-10	6.81E-08	9.64E-06	6.22E-04
EC_6.3.4.23	-0.611	1.40E-09	1.26E-07	1.70E-05	9.82E-04
EC_3.5.4.29	-0.591	1.52E-09	1.30E-07	5.68E-06	4.37E-04
EC_2.5.1.114	-0.585	1.90E-09	1.50E-07	2.17E-06	2.10E-04
EC_4.1.99.21	-0.536	1.94E-09	1.50E-07	9.39E-05	4.60E-03
EC_2.5.1.41	-0.571	2.13E-09	1.56E-07	3.55E-06	3.02E-04
EC_4.1.1.25	-0.563	2.37E-09	1.66E-07	2.26E-07	3.48E-05
EC_1.2.1.59	-0.62	2.95E-09	1.98E-07	4.38E-06	3.53E-04
EC_3.5.4.39	-0.597	3.29E-09	2.13E-07	2.19E-04	9.36E-03
EC_2.7.7.68	-0.534	3.48E-09	2.16E-07	1.30E-06	1.56E-04
EC_1.3.7.11	-0.534	1.02E-08	6.10E-07	1.01E-05	6.29E-04
EC_1.4.1.24	-0.52	1.60E-08	9.24E-07	1.06E-05	6.32E-04
EC_2.7.4.26	-0.53	1.67E-08	9.31E-07	8.25E-04	2.67E-02
EC_2.1.1.206	-0.54	2.04E-08	1.10E-06	2.93E-04	1.18E-02
EC_2.4.2.48	-0.472	5.53E-08	2.88E-06	8.21E-06	5.53E-04
EC_6.3.4.22	-0.486	1.47E-07	7.43E-06	5.22E-05	2.64E-03
EC_2.1.1.98	-0.477	2.32E-07	1.14E-05	3.48E-04	1.34E-02
EC_4.1.3.44	-0.505	2.64E-07	1.26E-05	1.68E-04	7.75E-03
EC_2.7.8.28	-0.472	5.97E-07	2.76E-05	3.20E-04	1.26E-02
EC_2.7.8.15	-0.426	9.09E-07	4.08E-05	2.32E-04	9.62E-03
EC_1.1.1.184	-0.232	2.96E-06	1.29E-04	2.20E-04	9.36E-03
EC_2.1.1.196	-0.391	6.44E-06	2.74E-04	4.59E-05	2.39E-03

EC_3.4.25.1	-0.521	1.77E-05	7.35E-04	1.88E-02	4.21E-01
EC_2.5.1.42	-0.362	2.63E-05	1.06E-03	8.22E-04	2.67E-02
EC_2.5.1.89	-0.364	3.30E-05	1.30E-03	3.10E-02	6.03E-01
EC_3.5.1.17	-0.194	3.54E-05	1.36E-03	5.31E-04	1.86E-02
EC_1.14.13.8	-0.167	6.08E-05	2.28E-03	1.15E-04	5.44E-03
EC_2.4.1.303	-0.143	6.35E-05	2.33E-03	5.73E-04	1.97E-02
EC_1.2.99.5	-0.709	1.47E-04	5.28E-03	1.74E-04	7.81E-03
EC_1.4.3.1	-0.281	2.16E-04	7.60E-03	6.20E-04	2.09E-02
EC_2.4.2.12	-0.271	2.75E-04	9.45E-03	3.83E-05	2.06E-03
EC_1.3.1.31	0.127	3.61E-04	1.22E-02	6.70E-03	1.72E-01
EC_2.7.1.7	-0.531	4.74E-04	1.56E-02	2.75E-08	6.34E-06
EC_1.1.1.30	-0.188	7.53E-04	2.43E-02	4.52E-04	1.62E-02
EC_1.20.1.1	-0.117	8.76E-04	2.77E-02	2.14E-03	6.52E-02
EC_1.7.2.5	-0.116	1.13E-03	3.52E-02	2.20E-01	9.96E-01
EC_1.1.1.261	-0.264	1.56E-03	4.77E-02	8.25E-03	2.05E-01
EC_2.3.2.10	-0.0964	2.21E-03	6.60E-02	1.96E-02	4.34E-01
EC_1.1.1.170	-0.165	2.44E-03	7.16E-02	6.53E-03	1.70E-01
EC_2.7.1.163	0.106	2.61E-03	7.52E-02	1.93E-03	6.00E-02
EC_4.2.1.120	-0.1	3.39E-03	9.61E-02	5.77E-01	9.96E-01
EC_3.1.3.13	-0.219	3.78E-03	1.05E-01	3.13E-03	9.19E-02
EC_6.3.5.2	-0.068	4.11E-03	1.13E-01	6.50E-03	1.70E-01
EC_1.3.1.26	-0.136	4.80E-03	1.29E-01	1.29E-03	4.08E-02
EC_3.1.4.4	0.235	6.68E-03	1.77E-01	5.11E-03	1.40E-01
EC_2.3.2.3	-0.118	7.57E-03	1.97E-01	3.60E-04	1.35E-02
EC_5.3.1.27	0.217	8.11E-03	2.08E-01	9.52E-03	2.33E-01
EC_2.7.8.25	0.182	8.71E-03	2.20E-01	6.38E-03	1.70E-01
EC_1.12.2.1	-0.213	1.19E-02	2.95E-01	4.41E-03	1.25E-01
EC_1.2.1.27	-0.157	1.24E-02	3.03E-01	2.42E-02	4.95E-01
EC_4.1.3.18	-0.163	1.37E-02	3.31E-01	4.95E-02	7.91E-01
EC_2.7.3.3	0.136	1.53E-02	3.63E-01	2.04E-02	4.39E-01
EC_2.1.1.195	0.192	1.55E-02	3.63E-01	9.95E-03	2.40E-01
EC_2.7.7.47	0.27	1.62E-02	3.69E-01	4.75E-03	1.32E-01
EC_2.8.3.11	-0.21	1.62E-02	3.69E-01	1.40E-02	3.22E-01
EC_2.7.1.36	-0.181	1.68E-02	3.77E-01	1.05E-02	2.50E-01
EC_3.5.1.81	0.125	1.78E-02	3.94E-01	8.10E-02	9.47E-01
EC_5.3.99.n1	0.0958	1.83E-02	3.99E-01	2.75E-03	8.22E-02
EC_3.2.1.96	0.324	1.86E-02	4.01E-01	2.50E-02	5.05E-01
EC_3.1.31.1	0.262	1.92E-02	4.03E-01	1.17E-02	2.73E-01
EC_6.3.2.39	0.226	1.92E-02	4.03E-01	5.97E-02	8.85E-01
EC_5.2.1.5	-0.12	2.07E-02	4.28E-01	7.75E-03	1.96E-01
EC_2.7.1.107	-0.34	2.30E-02	4.70E-01	1.81E-02	4.12E-01
EC_1.1.99.14	-0.0536	2.46E-02	4.96E-01	2.40E-02	4.95E-01
EC_4.1.2.43	0.124	2.71E-02	5.40E-01	4.31E-02	7.33E-01
EC_5.3.4.1	0.0712	2.82E-02	5.56E-01	4.23E-03	1.22E-01
EC_2.3.2.8	-0.307	3.04E-02	5.92E-01	5.40E-02	8.35E-01
EC_1.1.1.93	-0.25	3.10E-02	5.96E-01	2.18E-02	4.64E-01
EC_2.6.1.37	0.327	3.15E-02	5.98E-01	2.26E-02	4.73E-01

EC_1.2.2.2	-0.203	3.30E-02	6.15E-01	2.01E-01	9.96E-01
EC_1.3.99.3	0.0499	3.31E-02	6.15E-01	3.80E-02	7.05E-01
EC_5.1.1.4	0.0753	3.54E-02	6.50E-01	9.72E-02	9.96E-01
EC_4.2.1.106	0.123	3.78E-02	6.85E-01	2.96E-02	5.84E-01
EC_2.1.2.5	0.442	3.93E-02	7.04E-01	8.21E-02	9.47E-01
EC_3.1.4.17	-0.0612	4.00E-02	7.09E-01	7.66E-02	9.47E-01
EC_3.2.1.14	-0.219	4.38E-02	7.69E-01	3.86E-02	7.08E-01
EC_4.1.99.14	0.0654	4.46E-02	7.70E-01	7.05E-02	9.34E-01
EC_1.1.1.90	0.0559	4.48E-02	7.70E-01	1.41E-01	9.96E-01
EC_2.6.1.77	0.122	4.77E-02	8.06E-01	4.19E-02	7.33E-01
EC_3.11.1.1	0.33	4.84E-02	8.06E-01	3.60E-02	6.92E-01
EC_6.3.5.4	-0.206	4.84E-02	8.06E-01	1.03E-01	9.96E-01

Table S4. Enrichment analysis of metabolite classes. Z-scores and P-values of individual metabolites were from Wilcoxon rank-sum test comparing NAFLD$^{-/-}$ and NAFLD$^{-/+}$. Z-scores and P-values of metabolite classes were from Wilcoxon rank-sum test comparing the z-scores of metabolites in a metabolite class against the z-scores of metabolites in all other classes. Positive z-score indicates higher abundance in NAFLD$^{-/+}$, negative indicates higher abundance in NAFLD$^{-/-}$.

Metabolite class	number of metabolites	Metabolite class *P* value	Metabolite class z-score
Amino acids	24	0.017	2.38
Phenylpropanoic acids	7	0.063	-1.856
Pyridines	2	0.119	-1.56
Indoles	4	0.127	-1.526
Benzenoids	10	0.247	-1.157
Bile acids	18	0.431	0.787
Carbohydrates	2	0.535	0.62
Organic acids	22	0.57	-0.568
Fatty acids	34	0.769	0.294

Table S5. Correlations between metabolites and gut microbiota taxonomic profiles. Mantel tests (based on Spearman's correlation coefficients) were used to analyse the correlations of metabolites with thee taxonomic profiles at species, genus, and family level. Correlation coefficients and P-values are shown.

Metabolites	Species (corr. Coeff.)	Genus (corr. Coeff.)	Family (corr. Coeff.)	Species (*P* value)	Genus (*P* value)	Family (*P* value)
L-Lysine	0.015	0.037	0.059	0.3	0.084	0.017
L-Histidine	-0.005	0.038	0.024	0.55	0.068	0.171
Sarcosine	0.024	0.024	0.019	0.21	0.177	0.216
Beta-Alanine	0.006	-0.011	-0.026	0.42	0.614	0.802
L-Alanine	0.017	0.059	0.054	0.32	0.023	0.052
Gamma-Aminobutyric acid	0.062	0.014	0.003	0.05	0.313	0.463
L-Serine	0.032	0.009	-0.018	0.15	0.348	0.779
L-Homoserine	0.016	0.026	0.018	0.28	0.15	0.238
N-Acetylserotonin	-0.015	0.034	0.007	0.68	0.111	0.399
Hydroxypropionic acid	0.008	-0.001	-0.014	0.37	0.528	0.706
Ornithine	0.001	0.052	0.078	0.48	0.049	0.003
Nicotinic acid	0.03	0.013	0.025	0.11	0.254	0.126
2-Phenylglycine	0.026	0.041	0.041	0.25	0.107	0.081
L-Tyrosine	-0.007	0.002	-0.004	0.6	0.444	0.548
3-Chlorotyrosine	0.036	-0.01	-0.006	0.04	0.692	0.606
5-Hydroxy-L-tryptophan	0.033	-0.042	-0.037	0.18	0.904	0.87
L-Asparagine	-0.006	-0.008	-0.018	0.57	0.609	0.728
Hydroxyphenyllactic acid	0.097	0.045	0.042	0	0.048	0.06
L-Aspartic acid	0.017	-0.001	0.001	0.24	0.506	0.504
2-Hydroxy-3-methylbutyric acid	0.041	0.021	0.028	0.11	0.23	0.175
Picolinic acid	0.017	0.027	0.035	0.27	0.162	0.105
4-Hydroxybenzoic acid	0.012	-0.078	-0.057	0.39	0.995	0.969
Ortho-Hydroxyphenylacetic acid	0.056	-0.027	-0.035	0.05	0.838	0.87
Hippuric acid	-0.049	0.09	0.079	0.91	0.011	0.022
L-Malic acid	-0.042	-0.008	-0.023	0.91	0.626	0.777
Ethylmethylacetic acid	0.062	0.014	0.015	0.03	0.302	0.304
Benzoic acid	-0.031	-0.074	-0.074	0.82	0.995	0.995
2-Hydroxycaproic acid	-0.037	0.025	-0.001	0.83	0.244	0.501
Phenylacetic acid	0.042	0.066	0.075	0.12	0.015	0.009
2-Hydroxycinnamic acid	0.037	0.007	-0.008	0.12	0.398	0.606
Indoleacetic acid	0.009	0.104	0.107	0.39	0.002	0.001
Fumaric acid	0.024	0.022	0.028	0.21	0.182	0.145

Glutaric acid	0.03	-0.022	0.011	0.18	0.77	0.336
trans-Aconitic acid	-0.035	0.005	0	0.87	0.403	0.522
N-acetyltryptophan	0.014	0.03	0.028	0.31	0.146	0.16
Cis and trans-Cinnamic acid	-0.049	-0.022	0	0.96	0.754	0.497
2-Phenylpropionate	-0.066	0.049	0.052	0.98	0.056	0.063
Hydrocinnamic acid	-0.066	0.047	0.048	0.97	0.069	0.063
Indoleacrylic acid	-0.004	-0.045	-0.021	0.51	0.921	0.774
3-Indolepropionic acid	0.02	0.042	0.04	0.28	0.086	0.116
N-Phenylacetylphenylalanine	0.071	0.062	0.088	0.02	0.013	0.001
4-Hydroxyphenylpyruvic acid	0.021	-0.014	-0.011	0.24	0.694	0.63
Alpha-ketoisovaleric acid	0.074	0.049	0.029	0.01	0.045	0.144
3-Methyl-2-oxovaleric acid	0.053	0.061	0.036	0.03	0.008	0.065
Phenylpyruvic acid	0.049	0.012	-0.001	0.07	0.354	0.497
Methylmalonic acid	0.013	-0.037	-0.046	0.36	0.928	0.968
Glycine	0.062	0.022	-0.005	0.04	0.253	0.574
Glycolic acid	0.011	-0.01	-0.012	0.29	0.661	0.705
Glyceric acid	-0.026	0.041	0.015	0.76	0.088	0.313
L-Proline	-0.013	-0.011	-0.011	0.64	0.677	0.622
Formic acid	0.008	0.001	0.003	0.38	0.465	0.448
L-Valine	-0.018	-0.016	-0.017	0.73	0.76	0.761
Pyroglutamic acid	0.035	0.017	-0.002	0.12	0.25	0.549
3-Hydroxybutyric acid	-0.028	-0.053	-0.049	0.78	0.959	0.938
L-Methionine	0.013	0.031	0.051	0.31	0.125	0.035
alpha-Hydroxyisobutyric acid	0.058	0.018	0.031	0.04	0.234	0.105
2-Hydroxybutyric acid	-0.004	0.017	0.028	0.55	0.25	0.148
3-Hydroxyisovaleric acid	-0.02	-0.021	-0.033	0.73	0.786	0.898
Propanoic acid	0.035	0.018	0.029	0.12	0.24	0.134
Homovanillic acid	0.063	0.098	0.089	0.02	0.001	0.001
3- 3-Hydroxyphenyl -3-hydrox-ypropanoic acid	-0.029	0.019	-0.012	0.74	0.288	0.617
p-Hydroxyphenylacetic acid	0.035	0.077	0.06	0.13	0.009	0.032
L-Tryptophan	-0.023	0.002	-0.004	0.76	0.462	0.531
Butyric acid	0.027	0.019	0.022	0.16	0.216	0.186
Malonic acid	-0.009	0.079	0.073	0.58	0.014	0.02
Alpha-N-Phenylacetyl-L-gluta-mine	0.028	0.063	0.082	0.18	0.017	0.003
Tartaric acid	-0.035	-0.037	-0.024	0.85	0.885	0.769
Isovaleric acid	-0.018	0.089	0.088	0.7	0.006	0.004
Valeric acid	0.004	0.023	0.028	0.44	0.155	0.156

Phenyllactic acid	0.07	-0.035	-0.021	0.03	0.85	0.696
Oxalic acid	-0.012	0.08	0.064	0.6	0.007	0.017
Methylsuccinic acid	0.059	0.011	0.019	0.02	0.308	0.234
Adipic acid	0.065	-0.005	-0.001	0.06	0.555	0.507
Methylglutaric acid	0.021	-0.025	-0.02	0.21	0.866	0.823
cis-Aconitic acid	-0.01	0.002	0.003	0.61	0.465	0.447
Hydroxypyruvic acid	0.067	-0.018	-0.009	0.05	0.692	0.605
Suberic acid	0.017	0.022	0.03	0.34	0.269	0.193
Taurocholic acid	0.058	-0.003	-0.012	0.06	0.511	0.625
Citric acid	-0.035	-0.004	-0.001	0.9	0.569	0.481
Isocitric acid	-0.019	0.018	-0.008	0.75	0.235	0.596
2-Methylhexanoic acid	0.035	0.018	0.022	0.14	0.267	0.22
Heptanoic acid	0	0.013	0.022	0.48	0.307	0.215
Taurodeoxycholic acid	-0.021	-0.032	-0.017	0.71	0.851	0.737
Taurochenodeoxycholic acid	0.105	-0.027	-0.031	0	0.794	0.837
7-Dehydrocholic acid	0.043	-0.001	-0.002	0.1	0.494	0.519
12-Dehydrocholic acid	-0.027	-0.013	-0.01	0.96	0.816	0.764
w-muricholic acid	0.034	-0.008	-0.006	0.09	0.632	0.609
Oxoglutaric acid	0.012	0.006	-0.01	0.35	0.411	0.626
Glycohyocholic acid	0.028	-0.039	-0.023	0.21	0.889	0.771
Glycoursodeoxycholic acid	0.211	0.07	0.042	0	0.011	0.113
Octanoic acid	0.052	0.006	-0.011	0.06	0.42	0.609
muro-cholic acid	0.322	0.107	0.085	0	0.002	0.008
Ursodeoxycholic acid	0.17	0.033	0.029	0	0.154	0.183
Hyocholic acid	-0.025	-0.015	-0.009	0.93	0.812	0.717
Cholic acid	-0.031	0.007	0.025	0.82	0.43	0.211
Oxoadipic acid	-0.027	0.031	0.023	0.77	0.13	0.201
Glycocholic acid	0.013	0.003	-0.015	0.32	0.457	0.712
Nonanoic acid	0.028	-0.007	-0.015	0.19	0.605	0.695
Decanoic acid	0.02	0.011	0.007	0.27	0.333	0.369
Chenodeoxycholic acid	0.025	0.042	0.051	0.25	0.105	0.059
Glycochenodeoxycholic acid	0.081	0.009	-0.008	0.01	0.39	0.605
Glycodeoxycholic acid	0.038	0.001	0.011	0.14	0.476	0.376
Dodecanoic acid	0.07	-0.013	-0.029	0.03	0.656	0.866
Myristoleic acid	0.031	-0.001	-0.041	0.15	0.488	0.945
Lithocholic acid	0	0.078	0.092	0.49	0.015	0.007
Deoxycholic acid	0.024	0.068	0.092	0.25	0.022	0.002
Myristic acid	0.074	0.036	0	0.01	0.087	0.494

Pentadecanoic acid	0.063	0.046	0.009	0.03	0.052	0.354
Palmitoleic acid	0.031	0.003	-0.035	0.15	0.451	0.918
Alpha-Linolenic acid	0.043	0.021	0.001	0.07	0.191	0.457
Linoleic acid	-0.007	-0.004	-0.01	0.62	0.576	0.678
Arachidonic acid	0	0.013	-0.004	0.48	0.289	0.536
8,11,14-Eicosatrienoic acid	0.035	0.04	0.014	0.11	0.06	0.284
Docosahexaenoic acid DHA	0.032	-0.008	-0.037	0.17	0.619	0.921
Docosapentaenoic acid DPA	0.044	0.021	-0.02	0.09	0.221	0.766
Docosapentaenoic acid 22n-6	-0.032	-0.027	-0.024	0.89	0.892	0.847
Succinic acid	0.009	-0.047	-0.048	0.38	0.97	0.973
Citramalic acid	0.04	0.009	-0.006	0.12	0.387	0.556
Heptadecanoic acid	0.059	0.029	0.022	0.04	0.139	0.202
Stearic acid	0.047	0.029	0.01	0.05	0.147	0.341
Nonadecanoic acid	0.022	0.001	-0.01	0.21	0.47	0.652
11-cis-Eicosenoic acid	-0.025	-0.017	-0.028	0.76	0.739	0.821
3-Hydroxyphenylacetic acid	0.057	-0.025	-0.033	0.04	0.811	0.858

Table S6. SHAP-based inflection point analysis results using chi-square test of all 18 features in final model. Optimal thresholds of features and chi-square test P-values are shown.

Features	**Optimal threshold**	***P* value**
Phenyllactic acid	0.25	0.003
Hydrocinnamic acid	0.39	0.022
Methanobrevibacter	0.055	0.006
Fasting Insulin	6.22	0.018
L-Valine	286.13	0.049
Age	62.742	0.035
8,11,14-Eicosatrienoic.acid	51.5	0.051
Suberic acid	0.09	0.024
BMI	24.878	0.091
2-Phenylpropionate	0.29	0.037
HDL	1.4	0.034
N-Acetylserotonin	0.03	0.37
Oxoglutaric acid	70.34	0.041
N-Acetyltryptophan	0.24	0.018
PWY-3801	5.195	0.171
PWY-6167	1.412	0.033
PWY-7345	4.855	0.171
Slackia	0.023	0.097