eGastroenterology

Gut dysbiosis is linked to severe steatosis and enhances its diagnostic performance in MASLD

Marta Borges-Canha,[1,2] Javier Centelles-Lodeiro ,[3,4,5] Ana Rita Leite,[1,2] Joana Chaves ,[2] Inês Mariana Lourenço ,[2] Madalena Von-Hafe,[2] Catarina Vale,[2] Diana Martins ,[2] Cláudia Silva ,[2] António Carlos Ferreira,[2] Gwen Falony ,[5,6] Rodrigo Liberal ,[7] Mariana Fragão-Marques,[2] António Barros,[2] Isabel Miranda ,[2] Adelino Leite-Moreira,[2] Pedro Pimentel-Nunes,[8,9] Sara Vieira-Silva ,[5,10,11] João Sérgio Neves [1,2]

To cite: Borges-Canha M, Centelles-Lodeiro J, Leite AR, *et al*. Gut dysbiosis is linked to severe steatosis and enhances its diagnostic performance in MASLD. *eGastroenterology* 2025;**3**:e100204. doi:10.1136/egastro-2025-100204

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/egastro-2025-100204).

MB-C and JC-L are joint first authors.
SV-S and JSN are joint senior authors.

Received 24 February 2025
Accepted 21 July 2025

Linked

► https://doi.org/10.1136/egastro-2025-100196
► https://doi.org/10.1136/egastro-2024-100115
► https://doi.org/10.1136/egastro-2024-100086
► https://doi.org/10.1136/egastro-2023-100003

Check for updates

For numbered affiliations see end of article.

Correspondence to
Professor Sara Vieira-Silva; sara.vieira-silva@uni-mainz.de

ABSTRACT

Background Metabolic dysfunction-associated steatotic liver disease (MASLD) is the leading cause of chronic liver disease globally, with rising prevalence linked to metabolic syndrome (MetS). Excessive liver fat accumulation (steatosis) worsens disease progression and MASLD prognosis. Moreover, gut microbiota dysbiosis might promote steatosis, accelerating the disease progression to severe stages. Identifying gut microbiota signatures specific to steatosis severity might improve its diagnosis and inform personalised interventions in MASLD. This study aimed to characterise associations between gut microbiota composition and hepatic steatosis severity in a cohort of patients with MASLD/MetS. Ultimately, we aimed to assess the potential for microbiota features to enhance the diagnosis of severe steatosis.

Methods A cross-sectional cohort of 61 patients with MetS with extensive clinical history was recruited at different stages of MASLD. Transient elastography was used to evaluate liver fibrosis and steatosis severity. Participants' faecal microbiota were profiled using 16S rRNA gene sequencing. Statistical analyses first identified correlations between microbiota profiles and patients' phenotypes, while disentangling important confounders such as medication. Identified features were then used to build predictive models for diagnosing severe steatosis.

Results High steatosis severity was distinctly associated with a higher prevalence of the inflammation-associated Bacteroides 2 (Bact2)-enterotype, accompanied by a lower proportion of beneficial commensals (eg, *Akkermansia*) and a higher proportion of opportunistic bacteria (eg, *Streptococcus*). Patients harbouring a Bact2-enterotype reached severe steatosis at lower Fatty Liver Index (FLI) thresholds. Using Bact2-carrier status together with FLI in a predictive model significantly improved the classification of severe steatosis (accuracy 90%, receiver operating characteristics 96%) when compared with FLI alone.

Conclusion Gut microbiota composition and dysbiosis (defined as Bact2-enterotype) are distinctly associated with steatosis severity in MASLD/MetS. Patient stratification by microbiota composition enhances the diagnostic classification of severe steatosis in MASLD, suggesting a potential for personalised interventions in patients with microbiota dysbiosis.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Metabolic dysfunction-associated steatotic liver disease (MASLD) has become the most prevalent chronic liver disorder, driven by the global rise in metabolic syndrome (MetS). Hepatic steatosis exacerbates disease progression and worsens the prognosis of MASLD.
⇒ Gut microbiota dysbiosis is a hallmark of MASLD and MetS and might promote liver fat accumulation, accelerating the progression to severe steatosis. However, specific microbiota signatures independently associated with hepatic steatosis severity remain unclear.
⇒ Identifying such signatures might improve diagnostic performance, enabling earlier identification of patients with severe steatosis.

WHAT THIS STUDY ADDS

⇒ Our study identified a distinct disruption of the gut microbiota with increasing levels of steatosis, including an increased prevalence of the inflammation-associated enterotype (Bacteroides 2 (Bact2)-enterotype).
⇒ Incorporating Bact2-carrier status together with established clinical predictors (notably Fatty Liver Index) improved the diagnostic classification of severe steatosis.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Our findings support the potential for gut microbiota profiling and patient stratification by gut microbiota as a non-invasive solution to improve the diagnosis of severe steatosis in MASLD.
⇒ This could enhance early detection and help stratify patients at higher risk of severe hepatic conditions, thus improving MASLD management.
⇒ Future research is necessary to validate these microbial diagnosis markers and explore causal mechanisms for MASLD progression and prognosis in larger cohorts.

INTRODUCTION

Metabolic dysfunction-associated steatotic liver disease (MASLD), formerly known as non-alcoholic fatty liver disease (NAFLD),[1] has become the primary contributor to

chronic liver disease globally, approximately affecting 38% of the population.[2] The rapid increase in its prevalence is largely due to the parallel growth of metabolic syndrome (MetS).[3] MASLD is often referred to as the hepatic counterpart of MetS.[4 5] Metabolic disturbances in MetS, such as increased oxidative stress, chronic low-grade inflammation, insulin resistance and dyslipidaemia, contribute to excessive fat accumulation in the liver (steatosis), accelerating disease progression and worsening MASLD prognosis.[4–7]

Emerging evidence suggests that the gut microbiota might play a role in the onset and progression of hepatic steatosis in MASLD.[8–15] Dysbiosis, an imbalance in the gut microbial composition hindering its commensal role, is characteristic of MASLD and MetS. Patients with MASLD/MetS often show reduced diversity and beneficial commensals with an overrepresentation of opportunistic bacteria.[11–15] These imbalances might contribute to increased intestinal inflammation and gut permeability, leading to endotoxinaemia and low-grade systemic inflammation, which might promote steatosis and accelerate the progression to severe stages.[8]

Severe steatosis increases the risk of developing fibrosis, cirrhosis and hepatocellular carcinoma.[16] Therefore, effective management of steatosis is important for improving patient outcomes. Incorporating gut microbiota profiles alongside clinical diagnostic predictors of steatosis could enhance diagnostic performance for severe cases, allowing earlier identification of patients at high risk for severe steatosis and reducing the risk of progression to more severe stages, such as fibrosis.[11 12] However, specific gut microbiota signatures directly linked to hepatic steatosis remain incompletely understood.[16 17] Steatosis covariates such as obesity, medication use and other metabolic risk factors can impact gut microbiota signatures and therefore hinder the study of the relationship between the microbiota and steatosis.[9] Identifying gut microbiota signatures specific to steatosis—independent of confounding factors—might enhance the diagnosis of severe steatosis in MASLD.[15]

In this study, we aimed to examine the relationship between gut microbiota composition and hepatic steatosis in the microbiota e Doença Hepática Não Alcoólica (microDHNA) cross-sectional cohort, comprising patients with MetS across different stages of MASLD. We sought to determine whether steatosis was independently associated with distinct alterations in the gut ecosystem. Additionally, we aimed to evaluate whether incorporating gut microbiota profiles together with established clinical predictors could improve the diagnostic classification of severe steatosis.

METHODS

Study design and study participants

As previously described,[18] the microDHNA cohort recruited patients aged between 18 years and 75 years with a diagnosis of MetS, who were followed up in the outpatient setting of our tertiary centre. A total of 118 patients were screened between May 2021 and November 2022, and 65 patients were included in the cohort. Sixty-one participants provided a faecal sample and were included in the present analyses. The exclusion criterion was inability to provide informed consent. MetS was diagnosed by fulfilment of at least three of the five National Cholesterol Education Programme ATP III criteria.[19] The recruitment process included a detailed health questionnaire, physical examination, blood collection for predefined biochemical analyses, hepatic transient elastography and stool sample collection for microbiota analysis.

Clinical definitions

In agreement with clinical guidelines, we calculated the Fatty Liver Index (FLI) Score[20] as a diagnostic predictor of hepatic steatosis, and the body mass index (BMI), aspartate transaminase/alanine transaminase ratio, and diabetes (BARD),[21] the NAFLD Fibrosis Score (NFS)[22] and the Fibrosis-4[23] Index (FIB-4) scores as diagnostic predictors of hepatic fibrosis. The cut-off values FLI<30, BARD<2, FIB-4<1.3 and NFS<–1.46 were considered low-risk (clinically normal).[21]

Transient elastography

Liver stiffness (LS) and the controlled attenuation parameter (CAP) were measured using FibroScan (Echosens, Paris, France) to evaluate liver fibrosis and steatosis severity, respectively. An investigator with extensive experience (>1000 procedures) carried out all LS and CAP assessments using either the M or XL probe, in line with international guidelines and the manufacturer's instructions. The median LS value was recorded in kilopascals (kPa). According to international guidelines, only procedures with at least 10 valid readings, a success rate (ratio of number of successful measurements over total measurements) higher than 60%, and an interquartile range (IQR) for measurements/median ratio below 30% were included in the analysis.[24] CAP was assessed using radiofrequency data and the region of interest used for the LS measurements. The median CAP value was recorded in decibels per metre (dB/m). CAP and LS determination were evaluated simultaneously, and CAP values were only considered whenever LS measurements were valid. We applied commonly used cut-off values to define fibrosis and steatosis severity: ≤7.0 kPa for absent fibrosis, >7.0 kPa for the presence of liver fibrosis,[25] <236 dB/m for absent steatosis, ≥236 and <302 dB/m for mild-to-moderate steatosis, and ≥302 dB/m for severe steatosis[26] (table 1).

Stool sample collection and microbiota profiling

Stool samples were collected from each participant in their home environment. Samples were immediately stored in their home freezer. Frozen samples were delivered to our centre, processed and stored at −80°C until analysis. The time between sample collection and delivery did not exceed 24 hours. DNA was extracted from faecal samples

Table 1 Demographics and clinical characteristics of the microDHNA cohort

Variable	N	Summary
Age (years)	61	63 (57–70)
Gender (male)	61	43%
BMI (kg/m^2)	61	30 (27–32)
Absent steatosis (CAP <236 dB/m)	11	46.93%
Mild-to-moderate steatosis (CAP ≥236 dB/m and CAP <302 dB/m)	23	22.45%
Severe steatosis (CAP ≥302 dB/m)	15	30.61%
Waist-hip ratio	61	0.96 (0.91–1.00)
Glycated haemoglobin (A1c (%))	60	6.9 (6.0–7.7)
High-density lipoprotein (HDL (mg/dL))	60	49 (41–55)
Low-density lipoprotein (LDL (mg/dL))	60	87 (64–101)
Total cholesterol (mg/dL)	60	162 (131–178)
Triglycerides (mg/dL)	60	137 (86–172)
Estimated glomerular filtrate rate (eGFR (mL/min))	52	81 (70–98)
Albumin-creatinine (Cr) ratio	44	1 (1–1)
Steatosis (dB/m)	49	276 (239–322)
Fibrosis (kPa)	49	6.4 (4.2–6.6)
Aspartate transaminase (AST(U/L))	60	27 (19–30)
Alanine transaminase (ALT(U/L))	60	27 (17–30)
Gamma-glutamyl transferase (GGT(U/L))	58	46 (18–46)
Alkaline phosphatase (ALP (U/L))	54	75 (57–84)
Fatty Liver Index (FLI)	58	68 (51–88)
BMI AST/ALT Ratio and Diabetes Score (BARD)	60	3 (3–4)
Fibrosis-4 Index (FIB-4)	58	1.4 (0.9–1.6)
Non-alcoholic Fatty Liver Disease Fibrosis Score (NFS)	49	−0.92 (−1.90–-0.06)
Medication load	61	5.2 (4.0–7.0)
Diuretics	61	52%
Angiotensin receptor blocker (ARB)	61	49%
Beta blocker (BB)	61	28%
ACE inhibitor (ACEI)	61	34%
Calcium channel blocker (CCB)	61	28%
Mineralocorticoid antagonists	61	6.6%
Other antihypertensives (HTs)	61	1.6%
Sulphonylurea	61	15%
Acarbose	61	1.6%
Proprotein convertase subtilisin/kexin type nine inhibitor (PCSK9i)	61	3.3%
Clopidogrel	61	4.9%
Metformin	61	66%
Glucagon-like peptide-1 (GLP1) analogue	61	30%
Sodium-glucose co-transporter-2 inhibitor (SGLT2i)	61	31%
Dipeptidyl peptidase IV inhibitor (DPP4i)	61	16%
Insulin	61	51%
Statin	61	90%
Ezetimibe	61	15%
Fibrate	61	16%
Acetylsalicylic acid (ASA)	61	31%

Continued

Table 1 Continued

Variable	N	Summary
Hepatotoxic medication	61	39%
Alcohol (units)	61	41%
Coffee (units)	61	1.6 (1.0–2.5)
Fruit (units)	61	2.7 (2.0–3.5)
Yoghurt (units)	61	0.52 (0–1.00)
Exercise	61	52%

N (data available); Summary (percentage (%) or median (IQR)).
BMI, body mass index; CAP, controlled attenuation parameter; microDHNA, microbiota e Doença Hepática Não Alcoólica.

using the QIAamp PowerFecal Pro DNA Kit (Qiagen, Germany) following the manufacturer's instructions. Following extraction, DNA concentration was quantified using a Nanodrop One spectrophotometer (Thermo Scientific, Waltham, Massachusetts, USA) by measuring absorbance at 260 nm. The quality of the extracted DNA was assessed based on the 260/230 absorbance ratio and confirmed by agarose electrophoresis. DNA samples were sent to the Genomics Facility at Instituto Gulbenkian Ciência, where the V4 region of the 16S rRNA gene was sequenced using Illumina MiSeq technology.[27]

Computational methods

Metagenomic data preprocessing was performed using DADA2 V.1.29.0,[28] including filtering, trimming (filterAndTrim function with options: truncLen (130, 200), trimLeft (30, 30) and truncQ (11)), quality control, merging of sequencing pairs and taxonomy assignment with the Ribosomal Database Project classifier trainset V.18,[29] using the default parameters except otherwise specified. Samples were transformed using the variance stabilising transform, as recommended for compositional data,[30 31] with the poscounts option as fit type, ideal for zero-inflated data, and the local option for estimating dispersion.

Additional microbiota features were derived from the compositional profiles: microbiota alpha-diversity and community types (or enterotypes). Shannon's Diversity Index was calculated using the estimate richness function in *phyloseq* V.1.42.0.[32] Community-typing was performed using Dirichlet multinomial mixtures (DMM) models (dmn function in R) as described by Holmes *et al*[33] and Devolder *et al*,[34] using the genus-level phylogenetic count matrix (non-rarefied sequencing data). We combined the study cohort samples with 75 samples from alternative data sets (31 samples from, European Genome-phenome Archive (EGAC00001001168) and 44 samples from inhouse data sets from the same geographical region and with identical data generation handling). The DMM method applies probabilistic models to cluster samples with similar microbiota composition. The method is well suited for overdispersed compositional count data and provides biologically meaningful groupings. As strongly supported and consistently replicated in numerous Western population cohorts,[34–41] we set the optimal number of clusters to four. The probability for enterotype assignment was calculated for all samples, and the four clusters were labelled (for easier interpretation) after genera of predominant abundance in their respective model composition:[35] Rum (*Ruminococcaceae*; 31.1% of samples), Prev (*Prevotella;* 11.5% of samples), and Bact1 and Bact2 (*Bacteroides*; 14.8% and 42.6% of samples, respectively). For comparison with the current study, reference data on the population-level enterotype distribution (and its relation with BMI (kg/m^2)) from a Belgian cohort (the Flemish Gut Flora Project (FGFP), n=2345) were obtained from Vieira-Silva *et al*.[38]

Statistical analyses of the microbiota profiles were all performed in R statistical software (The R Foundation for Statistical Computing) using the packages vegan V.2.6–4,[42] *phyloseq*,*stats* V.4.2.2,[43] *FSA* V.0.9.4[44] and *rstatix* V.0.7.2.[45] Data visualisations were executed using the *ggplot2* V.3.4.2 R package.[46] All statistical tests were two-sided. Values of p were corrected for multiple testing using the Benjamini-Hochberg method (reported as adjP) when tests were performed on lists (n>1) of features for each data variable or data variable type and when performing multiple pairwise group (n>2) comparisons (eg, Kruskal-Wallis test with post hoc Dunn test). Significances were defined as p<0.05 and adjP<0.1. The Shannon Diversity Index was considered a stand-alone feature (n=1); raw values of p were reported. More details on specific statistical analyses are provided below.

Microbial community variation visualisation: Microbiota interindividual variation was visualised by unconstrained ordination principal coordinates analysis (PCoA) using Bray-Curtis dissimilarity on the genus-level compositional matrix (PCoA; ordinate function in R package *phyloseq*), with overlay of metadata associated with microbiota variation (displayed as arrows) using a post hoc fit on the PCoA (envfit function in R package *vegan*). The R package *loreplotr* V.0.2.1[47] was used to visualise changes in enterotype prevalence with continuous patient data (steatosis severity).

Correlation between patient metadata and microbial community variation: Constrained analysis of coordinates (single distance-based redundancy analysis (dbRDA);

capscale function in R package *vegan*; Bray-Curtis dissimilarity; genus-level) was used to quantify correlations between microbiota composition and single metadata variables. Metadata variables with a significant association (after correction for multiple testing by the Benjamini-Hochberg method) were then used in a stepwise dbRDA to obtain the best model, with maximum explanatory power (R^2) from metadata variables with non-redundant contributions (stepwise dbRDA; ordiR2step and capscale function in R package *vegan* (forward selection of variables); Bray-Curtis dissimilarity; genus-level). This analysis included variables summarised in table 1; medications taken by less than 10% of the cohort were excluded from further analyses.

Correlation between specific taxa and patient data: Taxa unclassified at the genus level or with less than 5 reads in 40% of samples were excluded from the association analyses. Spearman rank correlations were used for rank-order correlations between continuous variables. Wilcoxon rank-sum tests were used to test the median differences of continuous variables between two different groups, and Kruskal-Wallis tests with post hoc Dunn tests were used for more than two groups. One-sample Wilcoxon signed-rank tests were used to test whether median values of continuous variables were greater or smaller than a reference value. For categorical variables, statistical differences between patient groups were evaluated using pairwise χ^2 tests. Standardised effect sizes were retrieved for the abovementioned tests: ρ for Spearman rank correlations, Cramer's V for χ^2 tests, and rank biserial for one-sample Wilcoxon signed-rank tests and Wilcoxon rank-sum tests (*effectsize* V.0.8.3 R package).[48] Binary response and numerical response variables were respectively modelled using logistic regression (Logit; glm function in R package *stats*) and linear regression (lm function in R package *stats*), with single or multiple explanatory variables; numerical microbiota data were rank transformed to reduce skewness (rank function in the R package *stats*).

Prediction model building (classification of severe steatosis): We randomly stratified the microDHNA participants with measured steatosis severity (n=49) into a training set (n=39) and a holdout test set (n=10) isolated before training. The holdout test set consisted of a randomly selected balanced (equal representation) subset of participants with (n=5) and without (n=5) severe steatosis and was not used in model training. Prior to modelling, numerical microbiota data were rank transformed to reduce skewness (rank function in the R package *stats*) and all data were scaled and centred (across the training and holdout test sets) using the package *recipes* V.1.0.4[49] (step_scale and step_center functions). Missing FLI data (n=2 missing) were imputed with bagged tree models (step_impute_bag function). Bayesian logit models (bayesglm function in the *arm* V.1.13–1[50] R package) were trained on the training set (n=39) using package *caret* V.6.0–94[51], with repeated cross-validation (twofold cross-validation and 20 repeats). Because the training data set was imbalanced for severe steatosis status (Yes=25.6%; No=74.3%), upsampling was performed during model training to create a balanced data set with both categories having the same proportion (N max=58). Sensitivity, specificity, accuracy and the receiver operating characteristics (ROC) curve were calculated using the predict function from the *stats* package and the confusion matrix function from the *caret* package. Finally, the model's generalisability was evaluated on the holdout test set.

RESULTS

Clinical characterisation of the microDHNA cohort

We documented the anthropometric data of the microDHNA cross-sectional cohort. The cohort recruitment resulted in the inclusion of n=61 patients with MASLD/MetS, with an average age of 63 years (range 41–76), average BMI of 30 kg/m^2 (range 19–45) and a balanced gender distribution (57% female, 43% male). Patient characterisation also comprised liver function and structural metrics, medication and lifestyle parameters (table 1). All participants were medicated, taking between 2 and 10 different medications, with 59% experiencing polypharmacy (5 or more medications).[52] The cohort comprised patients across different stages of MASLD (online supplemental table S1 and figure 1), with a wide spectrum of steatosis (dB/m) progression (min=171 dB/m, max=400 dB/m). Steatosis severity and the FLI in the studied population significantly exceeded the clinically normal thresholds of 236 dB/m and 30, respectively (one-sample Wilcoxon signed-rank test, median=268, rank biserial=0.66, and median=71, rank biserial=0.95, respectively, $p<0.05$). Liver fibrosis (kPa) and the FIB-4 scores were within normal ranges of 7.0 kPa and 1.3, respectively (one-sample Wilcoxon signed-rank test, median=5.7 kPa, rank biserial=−0.47 and median=1.2, rank biserial=−0.054, respectively, $p>0.05$). However, the BARD and the NFSs (non-invasive predictors of advanced liver fibrosis) were higher than the clinical threshold of 2 and −1.46, respectively (one-sample Wilcoxon signed-rank test, median=3, rank biserial=0.91 and median=−0.71, rank biserial=0.43, respectively, $p<0.05$). In summary, the clinical characterisation of the patients revealed that this cohort was in the early stages of liver disease, as characterised by severe liver steatosis, but mild to moderate liver fibrosis (despite a clear BARD and NFS predicted risk).

Microbiota composition in the microDHNA cross-sectional cohort is linked to hepatic steatosis severity

To identify potential associations between the gut microbiota, a known contributor to human metabolic health, and liver function in MASLD, we explored which patient parameters most correlated with gut microbiota composition in our cohort (table 1, online supplemental table S2). Steatosis severity (continuous measure) and metformin intake emerged as significant, independent determinants of genus-level variation in the microbiota community

(stepwise multivariate dbRDA, full model R^2=4.34%; steatosis R^2=3.2%, adjP=0.004, metformin R^2=1.14%, adjP=0.046, N (multivariate model)=49; figure 1a). The impact of metformin on the gut microbiota composition is well documented and can be a confounder in microbiota studies, including medicated patients.[53–57] However, we confirmed that the association between steatosis and microbiota variation persisted when restricting the analysis to only metformin users (univariate dbRDA, R^2=3.54%; p=0.006, n=32). Moreover, steatosis-associated sequelae and risk factors—such as fibrosis and high BMI, respectively—could also potentially confound the steatosis-microbiota associations, but in our cohort, their correlation with microbiota composition was modest when tested individually (univariate dbRDA R^2: steatosis=3.2%, BMI=1.25%, fibrosis=0%) and non-significant when combined with steatosis (stepwise dbRDA, steatosis severity R^2=3.2%, adjP=0.006; BMI R^2=0.25%, adjP=0.33; fibrosis R^2=0.26%, adjP=0.33, n=49).

We identified four distinct enterotypes within our cohort labelled based on some of the dominant fractions of the clusters' representative composition,[35] termed Prev, Rum, Bact1 and Bact2 and with representative composition comparable to prior literature (online supplemental figure 2). Enterotypes have been linked to dietary preferences and to disease in Western populations.[34–36 38–41 58–60] In particular, the Bact2 enterotype has been found in multiple diseases to have the key characteristics of dysbiotic microbiota, including low density of microbial cells, low species diversity, depleted abundance of butyrate-producing bacteria and associated microbial metabolic pathways (eg, n-butyrate synthesis pathway), an essential mediator of host-microbiota homoeostasis.[34–41 58 59] Moreover, Bact2 has been found to be associated with systemic as well as gastrointestinal inflammation measured by C reactive protein (CRP) and faecal calprotectin levels, respectively.[34 38 40 41] We confirmed that in our cohort, Bact2 was associated with lowered diversity (Shannon Index) and decreased proportions of beneficial commensals such as *Faecalibacterium*, *Coprococcus*, *Oscillibacter*, *Roseburia*, etc (Kruskal-Wallis post hoc Dunn test, adjP<0.1, n=61, online supplemental figure 2 and table S3). Overall, 42.6% of the faecal samples from our cohort were classified as Bact2. This prevalence significantly exceeded the 13.4% observed in lean participants and the 19% observed in overweight or obese participants in the Western population, FGFP cohort (n=2345)[37 38] (χ^2 test, Cramer's V=0.17 and 0.21, respectively, adjP<0.001, figure 1b) (online supplemental table S4).

Having found that steatosis severity and metformin intake correlated with microbiota composition (figure 1a), we analysed their association with enterotypes (online supplemental table S5). We observed Bact2 to be more prevalent in patients with more severe steatosis (binomial logistic regression (Logit), OR=1.01, adjP=0.077, n=49, figure 1c). Metformin users also showed a higher prevalence of Bact2 (Logit, OR=5.20, adjP=0.040, n=61), as previously reported.[53] A similar increase in Bact2 prevalence in patients with more severe steatosis was observed in metformin users only (figure 1c), confirming our result was not confounded by this medication.

We next focused on identifying specific genera with abundances linked to steatosis severity and/or metformin use. The relative proportions of 15 genera were linked either with steatosis severity or with metformin, without overlap between the lists (online supplemental table S5). These distinct patterns suggest that steatosis and metformin intake have distinct impacts on the gut ecosystem. On one hand, steatosis severity was notably associated with decreased proportions of beneficial commensals such as *Akkermansia, Intestinimonas, Methanobrevibacter* and *Oscillibacter* and increased proportions of the opportunistic *Streptococcus* (Spearman rank correlation adjP<0.1, n=49, figure 1d,e). Metformin intake, on the other hand, was associated with increased proportions of the opportunistic genus *Escherichia* and reduced *Intestinibacter* (Wilcoxon rank-sum test adjP<0.1, n=61, figure 1d).

Enterotyping enhances the diagnostic classification of severe hepatic steatosis in MASLD

We next aimed to evaluate whether including microbiota profiles could enhance the diagnosis of steatosis beyond the predictive power of established clinical parameters. We first identified differences in microbiota features and clinical parameters between patients with divergent steatosis severity status. In 49 patients with measured steatosis severity, 15 (30.6%) were classified as having severe steatosis (CAP ≥302 dB/m) (see online supplemental table S6 for their comparative characteristics). Significant group differences in clinical parameters and gut microbiota features were identified (online supplemental table S7). Participants with severe steatosis had a higher BMI and FLI (Logit, OR>1, adjP<0.1, n=49, figure 2a), and a different microbial composition (univariate dbRDA, R^2=4.04%; p=0.001, n=49). These microbiota differences were also reflected—in the severe steatosis group—as a higher prevalence of the Bact2 enterotype (73.3% vs non-severe steatosis 29.41%) (Logit, OR=6.60, adjP=0.026, n=49, figure 2a), lower microbial diversity (Logit, OR=0.95, p=0.016, n=49, figure 2b), reduced proportions of *Akkermansia, Alistipes, Neglecta, Inhubacter* and *Sporobacter,* (Logit, OR<1, adjP<0.1, n=49, figure 2c) and increased proportions of *Streptococcus* (Logit, OR=1.06, adjP=0.07, n=49, figure 2c).

Building on these significant associations, we trained several predictive models to evaluate the added value of microbiota profiling in classifying severe steatosis. Recognising that microbiota profiling remains costly and time-consuming in clinical settings, we aimed to quantify whether integrating steatosis-associated microbial features in a prediction model could improve diagnostic performance beyond the performance achieved with well-established predictors of hepatic steatosis, such as the FLI. Therefore, we compared the classification performance of Bayesian Logit models trained with

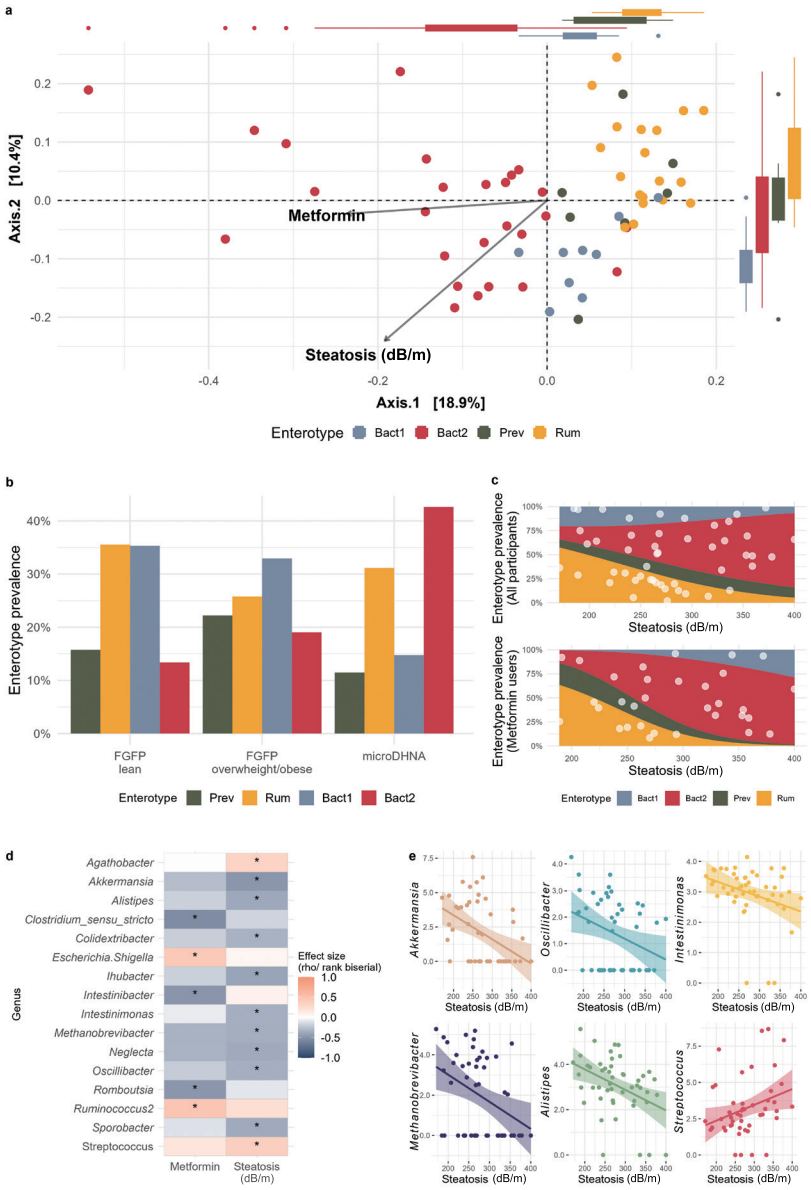

Figure 1 The gut microbiota composition in the microbiota e Doença Hepática Não Alcoólica (microDHNA) n=61 cohort is distinctly associated with steatosis severity (continuous measure, dB/m) and metformin use. (a) Visualisation of inter-individual differences in microbiota profiles (genus level composition) using principal component analysis (PCoA). Enterotypes (Bact2, Bact1, Rum, Prev) are represented by color. Vectors represent the post-hoc fit of the microbiota variation associated with the significant, independent determinants of genus-level variation in the microbiota community, steatosis severity and metformin intake (online supplemental table S2). Boxplots surrounding the PCoA represent the enterotype distribution data points along each PCo axis. The body of the box plot represents the first and third quartiles of the distribution, the line represents the median and the whiskers extend from the quartiles to the last data point within 1.5×interquartile range, with outliers beyond. (b) Enterotype distribution in the microDHNA cohort and compared with lean and overweight or obese participants in the Western population, Flemish Gut Flora Project (FGFP) data set. The microDHNA cohort showed greater prevalence of the dysbiotic Bact2 enterotype than lean and overweight/obese participants of the general Western population (n=2345, Chi-squared tests, online supplemental table S4). (c) Prevalence switch of enterotypes with steatosis severity in all participants and restricted to metformin users, showing the increase in Bact2 prevalence in patients with more severe steatosis (logistic regression (Logit), odds ratio=1.01, adjP=0.077, n=49, online supplemental table S5). Coloured areas represent the stacked enterotype prevalence along the steatosis gradient, with lines provided by multinomial Logit of enterotypes by steatosis severity, and data points (light grey) jittered at the corresponding steatosis severity level. (d) Genera proportions associated with steatosis severity and metformin intake (online supplemental table S5). Heatmap representation of the effect size of the associations (colour) (Spearman's rank correlation ρ for steatosis severity and Wilcoxon rank-sum test rank biserial for metformin). Adjustment for multiple testing was performed using the Benjamini-Hochberg method (adjP). The asterisks represent the significance level (adjP<0.1*). (e) Scatter plot representation highlighting genera proportions significantly correlated with steatosis severity (Spearman rank correlation test; adjP<0.1, online supplemental table S5). Bact1, *Bacteroides1*; Bact2, *Bacteroides2* Prev, *Prevotella*; Rum, *Ruminococcaceae*.

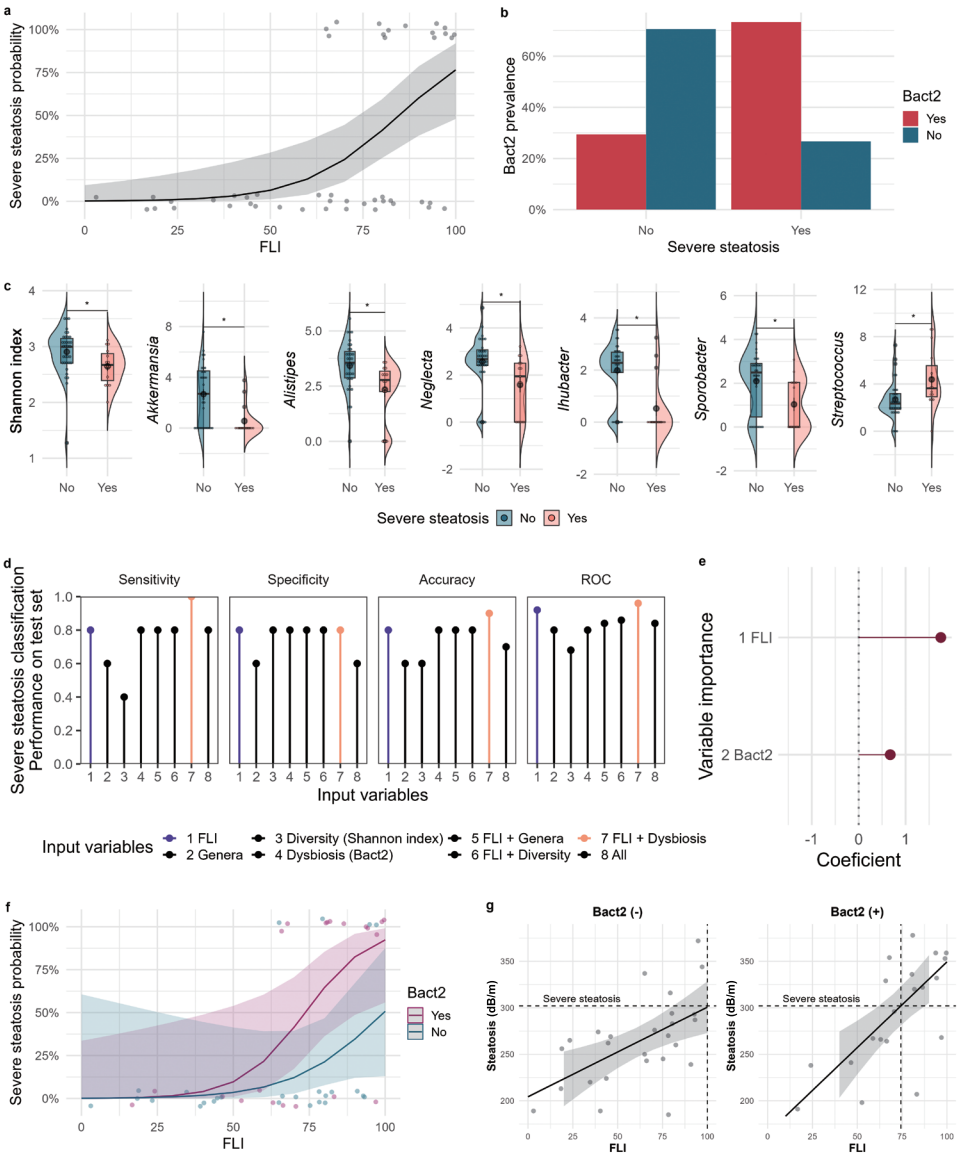

Figure 2 Performance of Fatty Liver Index (FLI) and microbiota profiling in classifying severe steatosis (≥302 dB/m) among patients with measured steatosis severity (n=49). (a, b, c) Higher FLI, positive Bact2 carrier status, lower diversity (Shannon index), a lower proportion of beneficial commensals and a higher proportion of opportunistic bacteria were associated with severe steatosis (online supplemental table S7). (a) Severe steatosis probability given FLI (logistic regression (Logit), FLI odds ratio (OR)=1.08, adjP=0.042, n=47). (b) Prevalence of Bact2 in participants with and without severe steatosis (Logit, Bact2 OR=6.6, adjP=0.026, n=49). (c) Box-violin plots highlighting microbial diversity and genera proportions significantly associated with severe steatosis (Logit, p<0.05 and adjP<0.1, respectively, n=49). Adjustment for multiple testing was performed using the Benjamini–Hochberg method (adjP). The asterisks represent the significance level (p<0.05* for Shannon index/adjP<0.1* for genera) (d) Bayesian Logit model performance for classifying severe steatosis in a holdout test set (online supplemental table S8). Model training was performed with cross-validation on a random selection of 39 of 49 patients (training set). Bayesian Logit models trained with the following variable sets: (1) FLI, (2) severe steatosis-associated genera, (3) microbiota diversity (Shannon index), (4) dysbiosis (Bact2 carrier status) and (5-8) combinations of the variable sets. The final evaluation of the models' performance (sensitivity, specificity, accuracy and receiver operating characteristic curve (ROC)) is shown in the panel and was executed on samples excluded from training (holdout test set, n=10). The holdout test set was balanced for severe steatosis status, including five patients with and five without severe steatosis (see methods and online supplemental figure 3). The FLI alone model performance is labelled in purple. The best-performing model (FLI+Dysbiosis), with the highest accuracy and ROC) in the holdout test set, is labelled in orange. (e) Variable importance in the best-performing model for severe steatosis (online supplemental table S8). The model included FLI (coefficient=1.75) and the Bact2 carrier status (coefficient=0.67). The variable importance of scaled variables is sorted from most important (top) to least important (bottom) and indicated by number (1 being the most important). (f) Severe steatosis probability given FLI plus Bact2 carrier status (multivariate Logit, n=47). (g) FLI threshold for severe steatosis in Bact2 carriers (+) (n=26) and non-carriers (+) (n=21) (online supplemental table S9). Bact2 carriers reached severe steatosis at a lower FLI threshold than non-carriers (74 vs 101, respectively) (linear regression, Bact2 carriers, coefficient=1.84, p<0.001; Bact2 non-carriers, coefficient=0.96, p<0.001). Bact2, *Bacteroides2*

the following variable sets: (1) FLI, (2) Severe steatosis-associated genera, (3) Microbiota diversity (Shannon Index), (4) Dysbiosis (Bact2-carrier status) and (5–8) Combinations of the variable sets. Model training was performed with cross-validation on a random selection of 39 of the 49 patients (training set). A final evaluation of models' performance was executed on samples excluded from training (holdout test set, n=10) to provide support for model generalisation (and rule out overfitting) (online supplemental figure 3).

The model, including FLI alone, had a high accuracy in classifying severe steatosis (80%) and ROC (92%) in the holdout test set. Among the microbiota features, Bact2 carrier status was the feature that had the best performance alone with an accuracy and ROC of 80%. The best-performing model was achieved by combining FLI and Bact2 carrier status, reaching an accuracy of 90% and an ROC of 96%, which corresponded to a 10% increase in accuracy and a 4% increase in ROC, as compared with FLI alone (online supplemental table S8, figure 2d). In this model, FLI remained the most important predictor (coefficient=1.75), followed by Bact2 carrier status (coefficient=0.67) (online supplemental table S8, figure 2e,f). Further analysing the relationship between FLI and severe steatosis in patients stratified by enterotypes (Bact2/not Bact2) revealed that Bact2 carriers reach severe steatosis at a lower FLI threshold than non-carriers (74 vs 101, respectively) (linear regression, Bact2 carriers, coefficient=1.84, p< 0.001, n=21; Bact2 non-carriers, coefficient=0.96, p=0.003, n=26, figure 2g, online supplemental table S9).

DISCUSSION

Recent research has explored the role of gut microbiota in the progression of MASLD. Patients with MASLD often show an imbalanced gut microbiota with reduced diversity and loss of beneficial commensals, and an overrepresentation of opportunistic bacteria.[11–15] Such imbalances might promote liver fat accumulation and accelerate the progression to severe steatosis.[8] Identifying gut microbiota signatures specific to steatosis severity, independent of confounding factors, is crucial for developing diagnostic predictors and possible new therapeutic strategies for MASLD in MetS. We recruited 61 participants with MetS across different stages of MASLD, who had minimal liver fibrosis but exhibited severe steatosis and steatosis-related scores. Our analyses of gut microbiota composition associated with steatosis severity revealed specific microbiota signatures linked to steatosis severity, independent of medication use (particularly metformin), and other potential confounders such as recognised risk factors (like high BMI). This signature was represented by decreased proportions of beneficial commensals such as *Akkermansia* and higher proportions of opportunistic bacteria such as *Streptococcus*, distinctly correlated with more severe hepatic steatosis. We also identified a novel association between steatosis severity and the prevalence of the Bact2 enterotype, previously described as being inflammation-associated in several diseases,[34 38 40 41] and notably characterised by a depletion of butyrate-producing bacteria, essential for host-gut microbiota homoeostasis.[34–41 58 59] Having explored both the microbiota signatures associated with steatosis severity (continuous measure) and with severe steatosis classification (patient classification on the clinical threshold of CAP ≥302 dB/m), we confirmed that the above-described signature also reflected the transition to severe steatosis.

We additionally explored how other patient clinical parameters were associated with microbiota composition. We could not find significant associations between microbiota composition and liver fibrosis in our cohort; this is likely due to limited variation in fibrosis severity among our recruited patients. Our observations do suggest that more severe steatosis is associated with a higher prevalence of dysbiosis (as defined by the Bact2 enterotype), before the progression of fibrosis. Regarding medication, our results support previous findings linking higher *Escherichia* abundance and lower *Intestinibacter* abundance with metformin use[54–57] as part of a potential side effect on microbiota composition. Indeed, a clinical trial identified similar microbiota alterations after metformin administration in treatment-naïve men without diabetes, thus disentangling the side effect from the therapeutic indication.[56] These microbiota alterations might contribute to the well-known gastrointestinal side effects of metformin, which is the main reason for treatment interruption.[55 61]

By building and comparing predictive models of severe steatosis, we provided supporting evidence that integrating microbiota features, particularly the Bact2 carrier status, together with FLI, improved the classification of severe steatosis. These results, as well as the fact that Bact2 carriers reached severe steatosis at a lower FLI threshold than non-carriers, suggest that gut microbiota dysbiosis might be a potential contributing risk factor for accelerated steatosis progression. The effective management of steatosis is essential for reducing the risk of MASLD progression in patients with MetS. Severe steatosis can lead to inflammation and subsequent fibrosis as a result of tissue repair. While microbial signatures in MASLD have been proposed previously as tools to differentiate between early and advanced stages of liver disease progression,[11–13] our results suggest that the microbiota could also facilitate differentiating between stages of liver steatosis in MASLD. While host metabolic and liver dysfunction, as reflected by the FLI, remained the strongest predictor of severe steatosis, incorporating gut microbiota profiles into clinical diagnostic predictors might offer additional clinical value, allowing for earlier diagnosis and closer management of patients with severe steatosis to prevent further complications such as fibrosis development.

In these predictive models, we also found that Bact2 carrier status classified severe steatosis better than microbial diversity or individual genera proportions. This is likely because community-typing captures dysbiosis while allowing for some variance in the composition and

might reflect a loss of function of the symbiotic microbiota. Indeed, the Bact2 enterotype has been shown to have limited functional capacity for butyrate production,[38] a key metabolite in the maintenance of immune tolerance and gut barrier integrity.[62] *In vitro* studies of faecal water from Bact2 carriers further revealed diminished levels of short-chain and medium-chain fatty acids and furan compounds, elevated bile acid concentrations and compromised epithelial barrier function.[58] Moreover, the Bact2 enterotype shows increased abundance of the genes involved in imidazole propionate biosynthesis, which might impair glucose metabolism and impact host inflammation.[59] These functional characteristics align with the Bact2 enterotype being associated with systemic as well as gastrointestinal inflammation, estimated, respectively, by CRP and faecal calprotectin levels.[34 38 40 41] While the current analysis, based on 16S rRNA gene sequencing, precludes microbial metabolism analysis, we did observe a depletion of butyrate producers. Future studies, including an assessment of metabolism through a multiomics approach, would be necessary to pinpoint microbiota-host metabolic interactions associated with MASLD, in particular, steatosis progression under different gut microbiota modulations.

STUDY LIMITATIONS

While our findings align with current evidence in microbiota disruptions associated with metabolic disorders, our cohort is of limited size and the generalisability of our results should therefore be confirmed with future experiments. Larger, multicentric cross-sectional cohorts would shed light on how much the Bact2 enterotype, other microbiota features or patient-specific factors might help as predictors of MASLD progression. A broader disease spectrum, inclusion of treatment-naïve patients, and diverse representation of clinical and lifestyle risk factors to MASLD would allow building more refined predictive models of disease stages.

CONCLUSION

In conclusion, our results show that a significant disruption in the gut microbiota is distinctly linked to steatosis severity. Incorporating gut microbiota profiles alongside well-established predictors for severe steatosis enhanced the diagnostic classification of severe steatosis, which could allow earlier identification of patients at high risk of developing fibrosis, cirrhosis and hepatocellular carcinoma.[16] Our results also suggested that dysbiosis could be contributing to steatosis progression, which might open new avenues for personalised interventions in patients with disrupted microbiota composition. Further studies—in larger cohorts with participants representing a broader range of liver disease, and more extensive phenotyping through multiomics data generation—are needed to validate and expand on our findings and deepen our understanding of host-microbiota interactions in this complex disease.

Author affiliations
[1]Serviço de Endocrinologia, Diabetes e Metabolismo, ULS São João, Porto, Portugal
[2]Cardiovascular R&D Centre—UnIC@RISE, Department of Surgery and Physiology, Faculty of Medicine, University of Porto, Porto, Portugal
[3]Department of Microbiology and Immunology, Rega Institute, KU Leuven, Laboratory of Molecular Bacteriology, Leuven, Belgium
[4]Vlaams Instituut voor Biotechnologie VIB-KU Leuven Center for Microbiology, Leuven, Belgium
[5]Institute of Medical Microbiology and Hygiene and Research Center for Immunotherapy (FZI), University Medical Center of the Johannes Gutenberg University Mainz, Mainz, Germany
[6]Host-Microbe Interactomics Group, Wageningen University, Wageningen, Netherlands
[7]Serviço de Gastrenterologia e Hepatologia, ULS São João, Porto, Portugal
[8]Department of Surgery and Physiology, University of Porto Faculty of Medicine, Porto, Portugal
[9]RISE@CI-IPO (Health Research Network), IPO-Porto, Porto, Portugal
[10]Institute of Molecular Biology (IMB), Mainz, Germany
[11]Institute for Quantitative and Computational Biosciences (IQCB), Johannes Gutenberg University Mainz, Mainz, Germany

Acknowledgements The authors thank all study participants for their valuable contribution to the study. GF and SV-S thank the support from the ReALity initiative at Johannes Gutenberg University and the Research Initiative Rhineland-Palatinate.

Contributors MB-C, JSN, JC, IML, MV-H, CV, ARL, AL-M and PP-N were responsible for the conceptualisation and design of the study cohort. JC-L, SV-S and GF are responsible for the conceptualisation and design of the statistical and modelling approach. MB-C, JC, IML, MV-H, CV, ARL, DM, CS, ACF, RL, MF-M, AB and IM conducted sample collection and data generation. JC-L performed all bioinformatic analyses and visualisation, supervised by GF and SV-S. JC-L, MB-C, GF, SV-S and JSN prepared the initial draft of the manuscript. All authors contributed to the interpretation of the results. The final version was revised and approved by all authors. Guarantor: SV-S. ChatGPT4o (OpenAI) was used to assist with grammar correction. The corrected grammar was always double-checked and no content was AI-generated.

Funding This study was funded by national funds through FCT - Portuguese Foundation for Science and Technology, under the scope of the Cardiovascular R&D Center – UnIC (UIDB/00051/2020 and UIDP/00051/2020). JC-L was supported by a travel grant fellowship from the Research Foundation Flanders (V465523N). SV-S is a principal investigator in the BMFTR Cluster4Future CurATime (CurATime project microbAIome; 03ZU1202CB) funded by the Federal Ministry of Research, Technology and Space (BMFTR). The funding agencies had no role in study design, analysis or interpretation, or writing of this article.

Competing interests SV-S and GF have received speaker fees from Ferring and Yakult, respectively. SV-S and GF are listed as inventors on patent WO2019115755A1 'A new inflammation-associated, low cell count enterotype', in the name of VIB VZW, Katholieke Universiteit Leuven, KU Leuven R&D and Vrije Universiteit Brussel. SV-S and GF are credited as inventors on WO2022073973A1 'Means and methods to diagnose gut flora dysbiosis and inflammation', in the name of VIB VZW, Katholieke Universiteit Leuven, KU Leuven R&D and University of Bristol. All other authors declare no competing interests.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by the Medical Ethics Committee of Centro Hospitalar Universitário de São João, Porto, Portugal (number 193/19). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data are available upon reasonable request. Raw microbiota sequencing data that support the findings of this study have been deposited in the European Nucleotide Archive under accession number PRJEB91737. Patient data will be made available upon reasonable request.

ORCID iDs
Javier Centelles-Lodeiro http://orcid.org/0000-0001-7454-5802
Joana Chaves http://orcid.org/0009-0004-2451-4556
Inês Mariana Lourenço http://orcid.org/0000-0003-0752-0802
Diana Martins http://orcid.org/0000-0003-4536-2315
Cláudia Silva http://orcid.org/0000-0003-2937-974X
Gwen Falony http://orcid.org/0000-0003-2450-0782
Rodrigo Liberal http://orcid.org/0000-0003-2525-1546
Isabel Miranda http://orcid.org/0000-0003-4839-1613
Sara Vieira-Silva http://orcid.org/0000-0002-4616-7602
João Sérgio Neves http://orcid.org/0000-0002-8173-8255

REFERENCES

1 Rinella ME, Lazarus JV, Ratziu V, *et al*. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. *Hepatology* 2023;78:1966–86.
2 Younossi ZM, Golabi P, Paik JM, *et al*. The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): a systematic review. *Hepatology* 2023;77:1335–47.
3 Younossi ZM, Koenig AB, Abdelatif D, *et al*. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* 2016;64:73–84.
4 Buzzetti E, Pinzani M, Tsochatzis EA. The multiple-hit pathogenesis of non-alcoholic fatty liver disease (NAFLD). *Metab Clin Exp* 2016;65:1038–48.
5 Marušić M, Paić M, Knobloch M, *et al*. NAFLD, Insulin Resistance, and Diabetes Mellitus Type 2. *Can J Gastroenterol Hepatol* 2021;2021:6613827.
6 Masenga SK, Kabwe LS, Chakulya M, *et al*. Mechanisms of Oxidative Stress in Metabolic Syndrome. *Int J Mol Sci* 2023;24:7898.
7 Monserrat-Mesquida M, Quetglas-Llabrés M, Capó X, *et al*. Metabolic Syndrome is Associated with Oxidative Stress and Proinflammatory State. *Antioxidants (Basel)* 2020;9:236.
8 Meroni M, Longo M, Paolini E, *et al*. A narrative review about cognitive impairment in Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD): Another matter to face through a holistic approach. *J Adv Res* 2025;68:231–40.
9 Vallianou NG, Kounatidis D, Psallida S, *et al*. NAFLD/MASLD and the Gut-Liver Axis: From Pathogenesis to Treatment Options. *Metabolites* 2024;14:366.
10 Moschen AR, Kaser S, Tilg H. Non-alcoholic steatohepatitis: a microbiota-driven disease. *Trends in Endocrinology & Metabolism* 2013;24:537–45.
11 Alferink LJM, Radjabzadeh D, Erler NS, *et al*. Microbiomics, Metabolomics, Predicted Metagenomics, and Hepatic Steatosis in a Population-Based Study of 1,355 Adults. *Hepatology* 2021;73:968–82.
12 Frost F, Kacprowski T, Rühlemann M, *et al*. Long-term instability of the intestinal microbiome is associated with metabolic liver disease, low microbiota diversity, diabetes mellitus and impaired exocrine pancreatic function. *Gut* 2021;70:522–30.
13 Loomba R, Seguritan V, Li W, *et al*. Gut Microbiome-Based Metagenomic Signature for Non-invasive Detection of Advanced Fibrosis in Human Nonalcoholic Fatty Liver Disease. *Cell Metab* 2017;25:1054–62.
14 Di Vincenzo F, Del Gaudio A, Petito V, *et al*. Gut microbiota, intestinal permeability, and systemic inflammation: a narrative review. *Intern Emerg Med* 2024;19:275–93.
15 Aron-Wisnewsky J, Vigliotti C, Witjes J, *et al*. Gut microbiota and human NAFLD: disentangling microbial signatures from metabolic disorders. *Nat Rev Gastroenterol Hepatol* 2020;17:279–97.
16 Natarajan Y, Kramer JR, Yu X, *et al*. Risk of Cirrhosis and Hepatocellular Cancer in Patients With NAFLD and Normal Liver Enzymes. *Hepatology* 2020;72:1242–52.
17 Sanyal AJ. Past, present and future perspectives in nonalcoholic fatty liver disease. *Nat Rev Gastroenterol Hepatol* 2019;16:377–86.
18 Borges-Canha M, Leite AR, Godinho T, *et al*. Association of metabolic syndrome components and NAFLD with quality of life: Insights from a cross-sectional study. *Prim Care Diabetes* 2024;18:196–201.
19 Grundy SM, Cleeman JI, Daniels SR, *et al*. Diagnosis and Management of the Metabolic Syndrome. *Circulation* 2005;112:2735–52.
20 Bedogni G, Bellentani S, Miglioli L, *et al*. The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population. *BMC Gastroenterol* 2006;6:33.
21 Cheah MC, McCullough AJ, Goh GBB. Current Modalities of Fibrosis Assessment in Non-alcoholic Fatty Liver Disease. *J Clin Transl Hepatol* 2017;5:261–71.
22 Angulo P, Hui JM, Marchesini G, *et al*. The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology* 2007;45:846–54.
23 Sterling RK, Lissen E, Clumeck N, *et al*. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology* 2006;43:1317–25.
24 European Association for Study of Liver, Asociacion Latinoamericana para el Estudio del Higado. EASL-ALEH Clinical Practice Guidelines: Non-invasive tests for evaluation of liver disease severity and prognosis. *J Hepatol* 2015;63:237–64.
25 Lin YS. Ultrasound Evaluation of Liver Fibrosis. *J Med Ultrasound* 2017;25:127–9.
26 Imajo K, Kessoku T, Honda Y, *et al*. Magnetic Resonance Imaging More Accurately Classifies Steatosis and Fibrosis in Patients With Nonalcoholic Fatty Liver Disease Than Transient Elastography. *Gastroenterology* 2016;150:626–37.
27 Walters W, Hyde ER, Berg-Lyons D, *et al*. Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *mSystems* 2015;1:00009–15.
28 Callahan BJ, McMurdie PJ, Rosen MJ, *et al*. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016;13:581–3.
29 Wang Q, Garrity GM, Tiedje JM, *et al*. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 2007;73:5261–7.
30 Lloréns-Rico V, Vieira-Silva S, Gonçalves PJ, *et al*. Benchmarking microbiome transformations favors experimental quantitative approaches to address compositionality and sampling depth biases. *Nat Commun* 2021;12:3562.
31 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.
32 McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* 2013;8:e61217.
33 Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS One* 2012;7:e30126.
34 Devolder L, Pauwels A, Van Remoortel A, *et al*. Gut microbiome composition is associated with long-term disability worsening in multiple sclerosis. *Gut Microbes* 2023;15:2180316.
35 Vandeputte D, Kathagen G, D'hoe K, *et al*. Quantitative microbiome profiling links gut community variation to microbial load. *Nature* 2017;551:507–11.
36 Valles-Colomer M, Falony G, Darzi Y, *et al*. The neuroactive potential of the human gut microbiota in quality of life and depression. *Nat Microbiol* 2019;4:623–32.
37 Falony G, Joossens M, Vieira-Silva S, *et al*. Population-level analysis of gut microbiome variation. *Science* 2016;352:560–4.
38 Vieira-Silva S, Falony G, Belda E, *et al*. Statin therapy is associated with lower prevalence of gut microbiota dysbiosis. *Nature* 2020;581:310–5.
39 Vandeputte D, Falony G, Vieira-Silva S, *et al*. Stool consistency is strongly associated with gut microbiota richness and composition, enterotypes and bacterial growth rates. *Gut* 2016;65:57–62.
40 Vieira-Silva S, Sabino J, Valles-Colomer M, *et al*. Quantitative microbiome profiling disentangles inflammation- and bile duct obstruction-associated microbiota alterations across PSC/IBD diagnoses. *Nat Microbiol* 2019;4:1826–31.

41 Caenepeel C, Falony G, Machiels K, *et al*. Dysbiosis and Associated Stool Features Improve Prediction of Response to Biological Therapy in Inflammatory Bowel Disease. *Gastroenterology* 2024;166:483–95.
42 Oksanen J, Simpson GL, Blanchet FG, *et al*. Vegan: community ecology package. Available: https://cran.r-project.org/web/packages/vegan/index.html [Accessed 25 May 2025].
43 R Core Team RCT. R core team (2013) R A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna; 2013. Available: http://www.R-project.org/ [Accessed 15 Nov 2025].
44 Ogle DH, Doll JC, Wheeler AP, *et al*. DunnTest()) AD (provided base functionality of. FSA: simple fisheries stock assessment methods; 2025. Available: https://cran.r-project.org/web/packages/FSA/index.html [Accessed 25 May 2025].
45 Kassambara A. Rstatix: pipe-friendly framework for basic statistical tests. 2023. Available: https://cran.r-project.org/web/packages/rstatix/index.html [Accessed 25 May 2025].
46 Wickham H. *Ggplot2: Elegant graphics for data analysis*. 2nd edn. Springer International Publishing, 2016.
47 Proost S, Vieira-Silva S, Raes J. Raeslab/lorepy: lorepy version 0.1.1. 2023.
48 Ben-Shachar M, Lüdecke D, Makowski D. effectsize: Estimation of Effect Size Indices and Standardized Parameters. *JOSS* 2020;5:2815.
49 Kuhn M, Wickham H. Preprocessing and feature engineering steps for modeling. 2023 Available: https://CRAN.R-project.org/package=recipes
50 Gelman A S, *et al*. Arm: data analysis using regression and multilevel/hierarchical models. 2024. Available: https://cran.r-project.org/web/packages/arm/index.html
51 Kuhn M. Building Predictive Models in R Using the caret Package. *J Stat Softw* 2008;28:1–26.
52 Masnoon N, Shakib S, Kalisch-Ellett L, *et al*. What is polypharmacy? A systematic review of definitions. *BMC Geriatr* 2017;17:230.
53 Forslund SK, Chakaroun R, Zimmermann-Kogadeeva M, *et al*. Combinatorial, additive and dose-dependent drug-microbiome associations. *Nature* 2021;600:500–5.
54 Forslund K, Hildebrand F, Nielsen T, *et al*. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* 2015;528:262–6.
55 Elbere I, Kalnina I, Silamikelis I, *et al*. Association of metformin administration with gut microbiome dysbiosis in healthy volunteers. *PLoS One* 2018;13:e0204317.
56 Bryrup T, Thomsen CW, Kern T, *et al*. Metformin-induced changes of the gut microbiota in healthy young men: results of a non-blinded, one-armed intervention study. *Diabetologia* 2019;62:1024–35.
57 Mueller NT, Differding MK, Zhang M, *et al*. Metformin Affects Gut Microbiome Composition and Function and Circulating Short-Chain Fatty Acids: A Randomized Trial. *Diabetes Care* 2021;44:1462–71.
58 Poppe J, Boesmans L, Vieira-Silva S, *et al*. Differential contributions of the gut microbiota and metabolome to pathomechanisms in ulcerative colitis: an *in vitro* analysis. *Gut Microbes* 2024;16:2424913.
59 Molinaro A, Bel Lassen P, Henricsson M, *et al*. Imidazole propionate is increased in diabetes and associated with dietary patterns and altered microbial ecology. *Nat Commun* 2020;11:5881.
60 Wu GD, Chen J, Hoffmann C, *et al*. Linking long-term dietary patterns with gut microbial enterotypes. *Science* 2011;334:105–8.
61 Bouchoucha M, Uzzan B, Cohen R. Metformin and digestive disorders. *Diabetes & Metabolism* 2011;37:90–6.
62 Louis P, Duncan SH, McCrae SI, *et al*. Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon. *J Bacteriol* 2004;186:2099–106.