Bacterial metabolic signatures in MASLD predicted through gene-centric studies in stool metagenomes

Juan Manuel Medina-Méndez[1,2], Paula Iruzubieta[2], Raúl Fernández-López[1], Javier Crespo[2], Fernando de la Cruz[1,*]

Author affiliations

[1] Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC), Universidad de Cantabria-CSIC, Santander, Spain

[2] Servicio de Gastroenterología y Hepatología. Grupo de Investigación Clínica y Traslacional en Enfermedades Digestivas. Instituto de Investigación Valdecilla (IDIVAL). Hospital Universitario Marqués de Valdecilla, Santander, Spain

[*] Corresponding author: delacruz@unican.es; Tel.: +34-942-201942

Abstract

Metabolic dysfunction-associated steatotic liver disease (MASLD) is a multifactorial condition influenced by the gut microbiome (GM). While previous studies have reported inconsistent associations between MASLD and key microbial clades using low-resolution 16S rRNA profiling, we employed high-resolution metagenomic sequencing and multi-marker taxonomic classification across three independent cohorts to identify robust microbial and functional signatures of MASLD. We consistently detected a depletion of *Agathobacter rectalis*, a known butyrate producer, in MASLD patients. Functionally, MASLD was characterized by a depletion of genes involved in butyrate and methane biosynthesis -particularly within the crotonyl-butyryl-CoA axis- alongside an enrichment of genes

driving the production of endogenous alcohols such as ethanol and 1-propanol. Genes encoding these fermentative pathways, often organized in operons like *pdu* and *tor*, were more abundant in MASLD samples, indicating a potential shift toward alcohol-producing metabolism. These geno-metabolic changes were accompanied by a broader displacement of beneficial taxa and an increase in accessory gene content across the GM, underscoring the limitations of taxonomy-based disease associations. Many of the differentially abundant genes were also found on plasmids, suggesting that horizontal gene transfer contributes to strain-level metabolic variability relevant to MASLD progression. Our findings support a model in which GM-driven metabolic shifts -rather than taxonomic changes alone- play a central role in MASLD pathogenesis, highlighting the importance of functional and mobile genetic element profiling for uncovering mechanistic links between the microbiome and liver disease.

Keywords

MASLD, gut microbiome, metagenomic sample, metabolic gene, RPKSM, differential gene abundance

1. Introduction

Metabolic-associated steatotic liver disease (MASLD) is the most common liver disease worldwide, affecting approximately one quarter of the global population [1]. Clinically, MASLD is defined by the presence of steatosis (i.e., accumulation of intrahepatic triglycerides) in more than 5% of hepatocytes in association with metabolic risk factors (particularly, obesity and type 2 diabetes) and in the absence of excessive alcohol consumption (≥30 g per day for men and ≥20 g per day for women) or other chronic liver diseases [2]. However, the term encompasses a spectrum of pathological situations, ranging from simple steatosis to non-alcoholic steatohepatitis (NASH), which can further progress to fibrosis, cirrhosis and hepatocellular carcinoma [3]. MASLD also significantly increases the

42 risk of cardiovascular disease, extrahepatic malignancies (e.g., colorectal or breast cancer), liver-related
43 complications or chronic kidney disease [4]. The pathogenesis of MASLD is multifactorial, involving
44 genetic, environmental, inflammatory and metabolic factors, and it has a bidirectional association with
45 components of the metabolic syndrome, particularly insulin resistance [5].

46 The gut microbiome (GM) interacts with the liver via the gut-liver axis (GLA) through the portal
47 vein, which transports GM-derived metabolites towards the liver. This axis can be disrupted by multiple
48 interconnected factors, affecting the progression of MASLD and fibrosis [6]. An alteration of the GM
49 composition and/or their associated metabolic production can lead to an increased production of
50 harmful metabolites such as pathogen-associated molecular patterns, which induce inflammatory
51 responses in the liver when they reach it through the bloodstream, leading to liver injury [7, 8]. The
52 transient of these endotoxins into portal circulation is facilitated by an increased intestinal permeability
53 caused by a disruption of tight junction proteins in the intestinal membrane [9]. Also, an alteration of
54 the bile acid composition and circulation affects liver metabolism and function. Because they play a role
55 in the regulation of glucose and lipid metabolism, their dysregulation contributes to gut inflammation
56 and MASLD [10, 11].

57 Several studies have tried to establish a causal link between bacterial metabolism and MASLD
58 [12]. However, metabolic levels can also be influenced by host metabolism, diet or environmental
59 factors. Similarly, metataxonomic analysis have explored the relationship between GM composition
60 and MASLD [13, 14]. Although 16S rRNA gene sequencing provides taxonomic resolution at the genus
61 or species level, it fails to capture strain-level metabolic variability. Bacterial genomic plasticity allows
62 key metabolic genes, such as those potentially relevant to MASLD pathogenesis, to be encoded in their

63 accessory genome, including mobile genetic elements, further complicating the association between
64 taxonomy and metabolic function [15]. In this study, we used genes encoding the metabolic enzymes
65 involved in the final steps of microbial pathways leading to the production of butyrate, ethanol,
66 propanol and methane as proxies to determine their abundance in MASLD. To achieve this, we analyzed
67 their abundance in over 550 whole metagenomes obtained from fecal samples of MASLD and healthy
68 patients, stratified across three different cohorts.

69 **2. Results**

70 We retrieved 554 metagenomic sequencing libraries from the human fecal samples collected in
71 five different studies and stored in the European Nucleotide Archive (ENA) [16]. These samples were
72 organized in three MASLD cohorts, each split into two phenotypic groups of comparison, as detailed in
73 Table 1.

Cohort	Phenotypic group	N samples	Reference	Accession number
1	Healthy control	117	[17]	PRJNA544527
	MASLD, initial fibrosis (F0-F2)	73	[18]	PRJEB14215
		48	[19]	PRJNA420817
2	MASLD, initial fibrosis (F0-F2)	72	[20]	PRJNA373901
	MASLD, advanced fibrosis (F3-F4)	14		
3	Healthy control	105	[21]	PRJEB6337
	Cirrhosis MASLD	125		

74 **Table 1. MASLD patient cohorts analyzed in this study.** Each cohort (column 1) was divided into two phenotypic comparison
75 groups: control vs. disease (column 2). The cohorts include N MASLD patients and corresponding whole-metagenomic stool
76 samples derived from them (column 3). The disease group in Cohort 1 consists of 73 and 48 MASLD patients with initial
77 fibrosis from [18] and [19], respectively. Cohort 2 was divided into initial fibrosis (control) and advanced fibrosis. Details on
78 cohort organization are provided in the Materials and Methods section. Patients were originally recruited in the listed
79 studies (column 4), with metagenomic data and medical metadata downloaded from the indicated ENA accession numbers
80 (column 5).

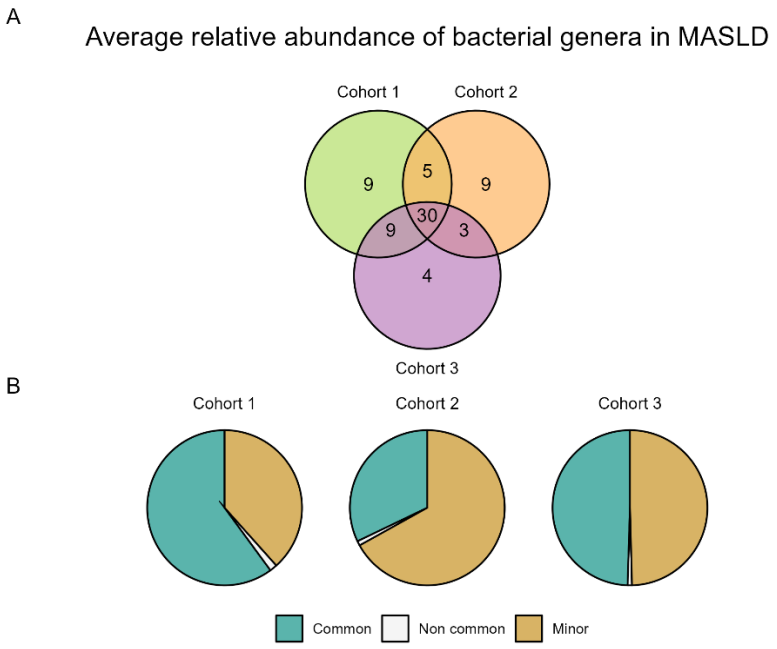
81 **Bacterial taxonomic signatures are inaccurate MASLD predictors**

82 Certain bacterial clades have been positively correlated with MASLD progression, including
83 *Enterobacteriaceae* [14, 22] such as *Escherichia / Shigella spp*. [20, 22], *Bacteroidaceae* like *Bacteroides*
84 *spp.* [20, 23], *Veillonellaceae* [14] or *Streptococcus spp.* [24]. Conversely, other clades have been
85 reported to exhibit an inverse correlation with MASLD progression, including *Ruminococcaceae* [14]
86 such as *Faecalibacterium prausnitzii* [20, 25], *Lachnospiraceae* like *Agathobacter rectalis* [20, 25] or
87 *Dorea longicatena* [25], *Clostridiaceae* like *Clostridium spp.* [22], or even *Pseudomonas spp.* [24].
88 However, these studies often yield conflicting results. For example, while [23] reported a positive
89 correlation between *Ruminoccocus spp.* abundance and MASLD progression, this finding contradicts
90 the one of the main findings in [20, 25]. Similarly, *Prevotella spp.* was found to decrease with MASLD
91 development [26] and fibrosis stage [23], yet [24] reported an increase in this genus in MASLD patients.
92 Additional examples of discordant taxonomic signatures in MASLD, such as *Blautia spp.* or *Roseburia*
93 *spp.,* are reviewed in [27].

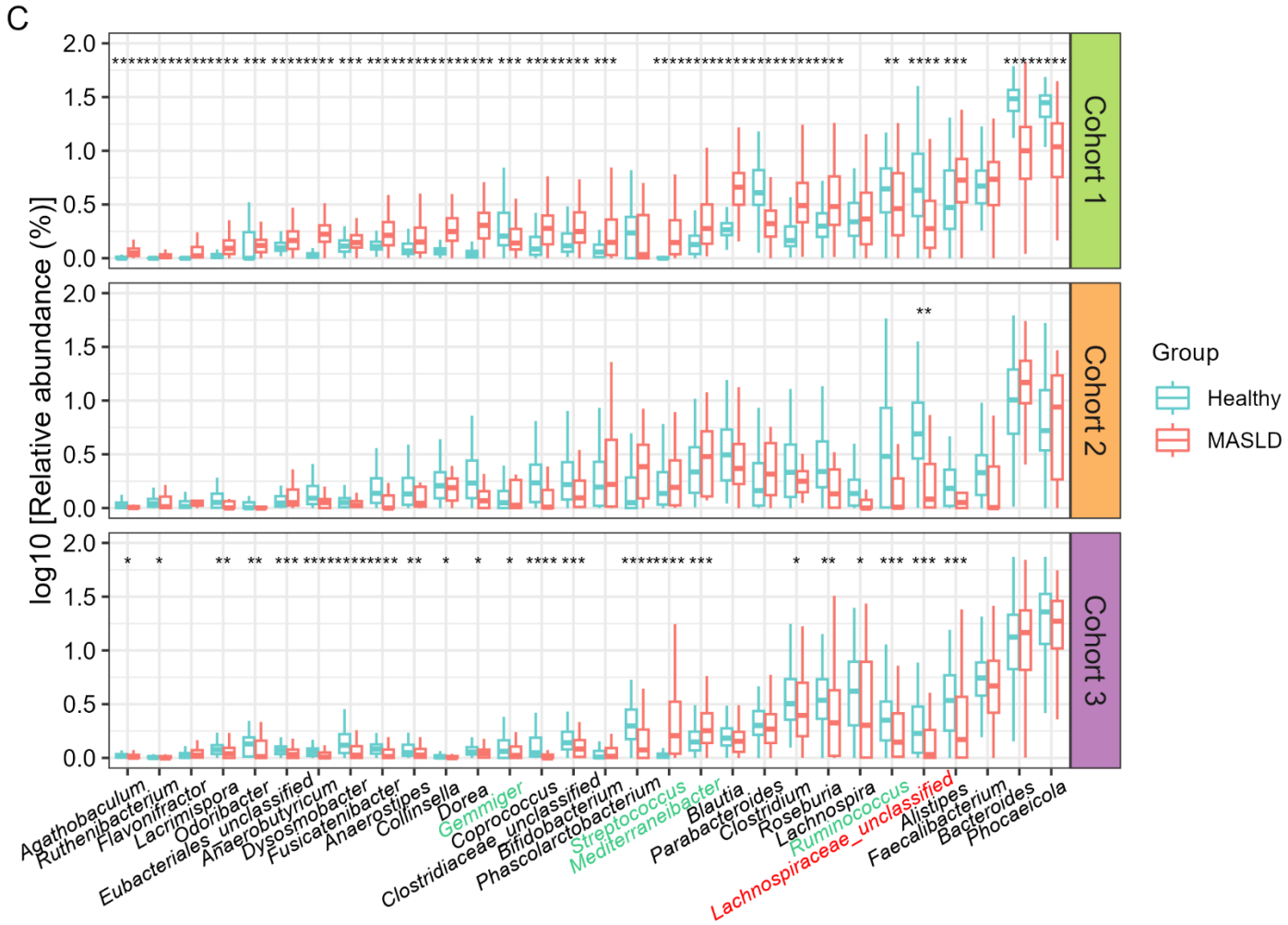
94 To investigate whether any bacterial genera is systematically altered in MASLD, we profiled the
95 taxonomic composition of the GM communities across all samples from the three cohorts using
96 MetaPhlAn [28, 29]. Unclassified and low abundant genera (i.e., those with an average relative
97 abundance of 0) were excluded, as well as genera present in only one group, as detailed in Materials
98 and Methods. This analysis identified 30 bacterial genera shared across all three cohorts (Fig. 1A) that
99 exhibited significant differences in relative abundance between conditions in at least one cohort (Fig.
100 1C). The presence of differentially-abundant genera exclusively in one cohort could be attributed to
101 heterogeneous demographic variables such as geographical location, age, sex, or ethnicity, as well as
102 environmental factors like diet, physical activity, comorbidities or drug exposure [30, 31], rather than

being directly related to the MASLD phenotype. These 30 genera accounted for 60%, 32% and 49,5% of the bacterial fraction in each cohort, respectively (Fig. 1B). Thus, nearly half of the bacterial genera identified in at least one cohort are shared across all cohorts (30 out of 69), comprising between one-third and two-thirds of their total gut bacterial abundance. Among these 30 shared genera, only *Gemmiger, Streptococcus, Mediterraneibacter, Ruminococcus* and an unclassified *Lachnospiraceae* exhibited a consistent differential abundance pattern between across every cohort. However, only the unclassified *Lachnospiraceae* showed statistically significant differences between the groups (Fig. 1C). These findings suggest that no bacterial genera are consistently and robustly altered in MASLD.

Figure 1 (Next page). Abundance of GM bacterial genera in MASLD. (A) Number of bacterial genera common to the three cohorts. **(B)** Average relative abundance of bacterial genera predicted by MetaPhlAn in the three cohorts of study. See Table 1 for details of the group composition from each cohort. The blue section of the pie charts reflects the average relative abundance of bacterial genera common to the three cohorts (n=30 genera), whereas the grey and green sections reflect, respectively, the abundances of non-common and minor genera (i.e., with an average relative abundance = 0 in at least one of the cohorts). **(C)** Relative abundance of bacterial genera in the metagenomes of the fecal samples, in logarithmic scale. Each facet represents a cohort of study. Differences in abundances were evaluated using a pairwise Mann-Whitney test with a Benjamini-Hochberg adjustment for multiple testing. Horizontal bars represent median levels, while boxes and whiskers represent the data from first to third quartiles and from the quartiles to the minimum and maximum, respectively. Genera depicted in green across the x-axis have a consistent differential abundance pattern between groups over the three cohorts. Only genus *Lachnospiraceae_unclassified*, depicted in red, has a consistent and statistically significant increase between groups across cohorts. (*) $p<0.05$, (**), $p<0.01$, (***), $p<0.001$, (****) $p<0.0001$. Non-starred genera present non-significant differences in abundance between groups.


A
Average relative abundance of bacterial genera in MASLD
Cohort 1
Cohort 2
Cohort 3
9
5
9
30
9
3
4
B
Cohort 1
Cohort 2
Cohort 3
Common
Non common
Minor

126


C
log10 [Relative abundance (%)]
2.0
1.5
1.0
0.5
0.0
Cohort 1
Cohort 2
Cohort 3
Group
Healthy
MASLD
Agathobaculum
Ruthenibacterium
Flavonifractor
Lacrimispora
Odoribacter
Eubacteriales_unclassified
Anaerobutyricum
Dysosmobacter
Fusicatenibacter
Anaerostipes
Collinsella
Dorea
Gemmiger
Coprococcus
Clostridiaceae_unclassified
Bifidobacterium
Phascolarctobacterium
Streptococcus
Mediterraneibacter
Blautia
Parabacteroides
Clostridium
Roseburia
Lachnospira
Ruminococcus
Lachnospiraceae_unclassified
Alistipes
Faecalibacterium
Bacteroides
Phocaeicola

127

128 **Bacterial genes involved in the formation of butyrate are depleted in MASLD**

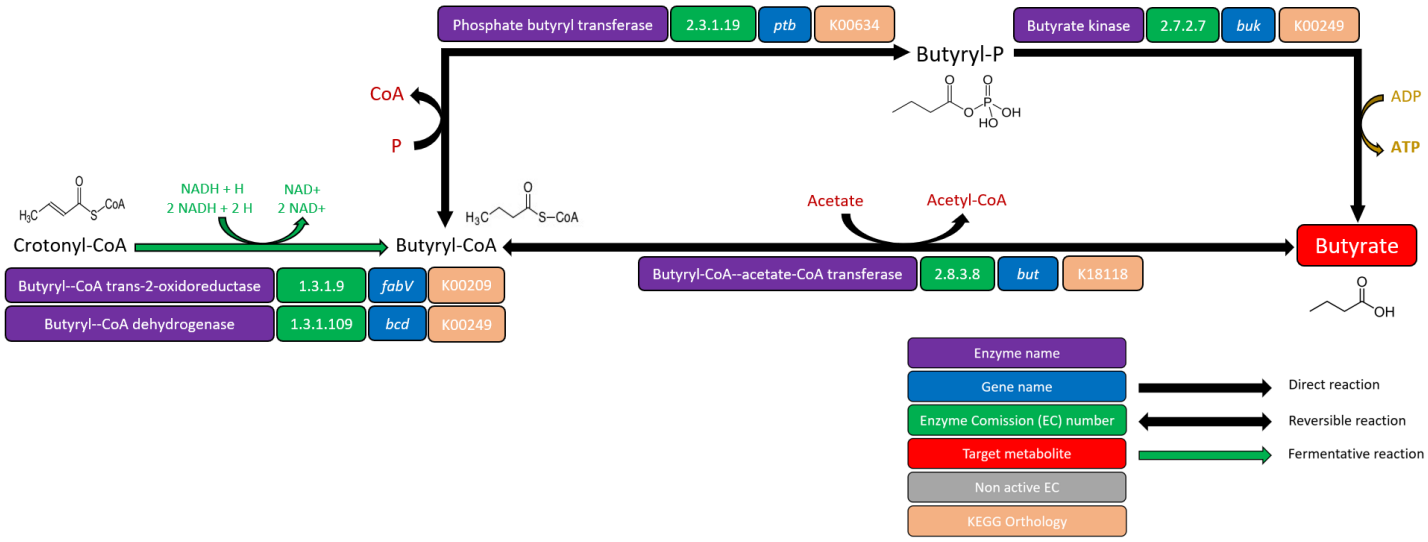
129 The absence of consistent taxonomic signatures across multiple MASLD cohorts does not
130 exclude a potential role for the GM in MASLD development. Certain microbial metabolites regulate GLA
131 integrity and have been suggested to influence MASLD progression [32–34]. These compounds are
132 catabolized from dietary carbohydrates and proteins that escape upper gastrointestinal digestion by
133 the GM through fermentative pathways that are predominant in the anaerobic environment of the
134 colon [35, 36]. Once absorbed into the bloodstream, they are transported to the liver via the portal
135 vein, where they can exert both beneficial and detrimental effects [37, 38]. For this reason, we
136 interrogated the metagenomic fecal samples from the patient cohorts for metabolic signatures that
137 could explain the MASLD phenotype. First, we focused on the genes encoding enzymes involved in
138 butyrate formation as a proxy for the bacteria that carry them, independently of their taxa. To validate
139 our hypothesis, we analyzed the genic abundance of the final two enzymes in butyrate production. We
140 selected butyrate as a target because it is crucial for maintaining GLA homeostasis by supporting the
141 colonic epithelial barrier [39, 40]. It also promotes lipid oxidation [41, 42], induces hepatic
142 gluconeogenesis [43], and prevents insulin resistance [44], all of which are key factors in MASLD. In
143 mice, both butyrate-producing probiotics [45] and sodium butyrate administration [46–48] have been
144 shown to prevent steatosis and slow MASLD progression. Additionally, fecal microbiota transplantation
145 (FMT) from healthy to NASH mice increased cecal butyrate levels and reduced steatosis [49].

146 Bacterial butyrate can be formed through two different terminal reactions (Fig. 2A). The first is
147 a dephosphorylation of butyryl-P that yields one ATP through an energetically favorable reaction that
148 is catalyzed by a butyrate kinase encoded by *buk.* In the precursor reaction, butyryl-P is formed through
149 a phosphorylation of butyryl-CoA catalyzed by a phosphate butyryl transferase encoded by *ptb*.

150 Butyrate also accumulates after the transference of a CoA cofactor from butyryl-CoA to acetate through
151 butyryl-CoA:acetate CoA transferase, encoded by *but*. This CoA-transferase route also conserves the
152 energy of the CoA bond in the newly formed CoA-moiety of the co-substrate. Butyryl-CoA is produced
153 from crotonyl-CoA through two different enzymes: butyryl--CoA trans-2-oxidorreductase, encoded by
154 *fabV*, and butyryl--CoA dehydrogenase, encoded by *bcd.* This pathway is further reviewed in the human
155 GM by [50].

156 We quantified the abundance of these genes in the metagenomic fecal samples of the two
157 groups across the three MASLD cohorts, as described in Materials and Methods. Genes involved in the
158 production of butyrate and its precursors, butyryl-P and butyryl-CoA, were significantly depleted in the
159 GM of cirrhosis MASLD patients compared to the healthy group (Fig. 2B, bottom), and they also
160 decreased significantly with MASLD progression (Fig. 2B, up). Precursor gene *fabV* was nearly an order
161 of magnitude less abundant than *ptb* and *bcd* in both cohorts, underscoring the greater prevalence of
162 *bcd* over *fabV* in the CoA-transferase route. Additionally, *ptb* did not exhibit significant differences in
163 Cohort 2. *but* was significantly depleted in the control group in both cohorts, whereas *buk* showed no
164 differences in Cohort 2 either. Notably, the depletion of butyrate genes did not replicate in MASLD
165 patients from Cohort 1.

2A)



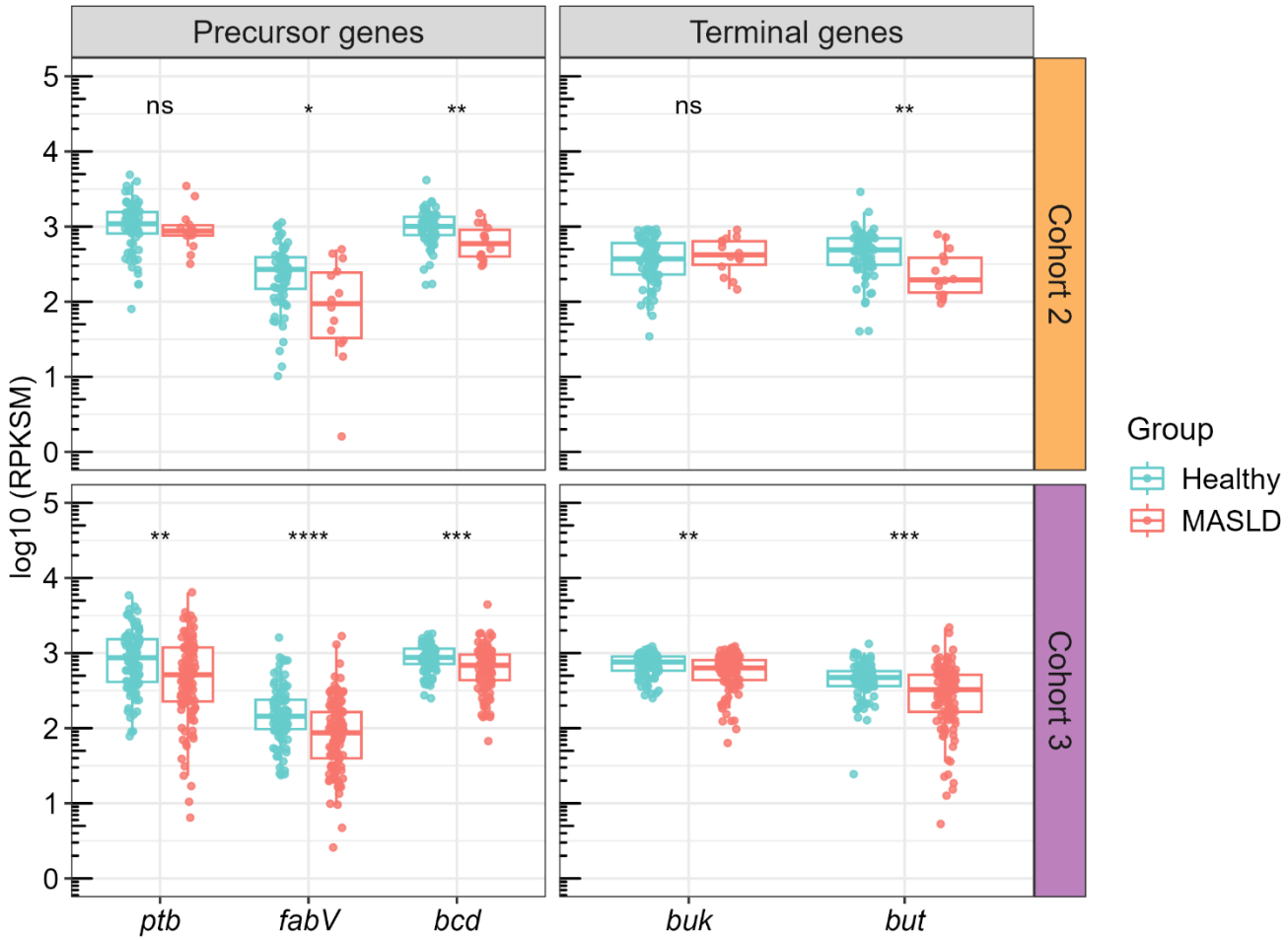


Figure 2. Butyrate-producing genes are depleted in MASLD. (A) Characterization of the terminal reactions involved in the formation of bacterial butyrate from crotonyl-CoA. Enzymes, EC numbers, coding genes and KOs, as well as the reaction type, are indicated for each metabolic step according to the annexed legend. **(B)** Abundance of butyrate-producing genes in Cohorts 2 and 3, in logarithmic scale. Boxplots represent the genic abundance from the metagenomic measures of the

172 fecal samples, and differences in abundances were evaluated using a pairwise Mann-Whitney test with a Benjamini-
173 Hochberg adjustment for multiple testing. Upper and lower facets represent abundances in Cohorts 2 and 3, respectively.
174 See Table 1 for details of the group composition from each cohort. Left and right facets represent abundances of precursor
175 and terminal coding genes, respectively, as described in (A). Horizontal bars represent median levels, while boxes and
176 whiskers represent the data from first to third quartiles and from the quartiles to the minimum and maximum, respectively.
177 RPKSM: reads per kilobase per genic size per million reads. (*) $p<0.05$, (**), $p<0.01$, (***), $p<0.001$, (****) $p<0.0001$, ns:
178 non-significant.

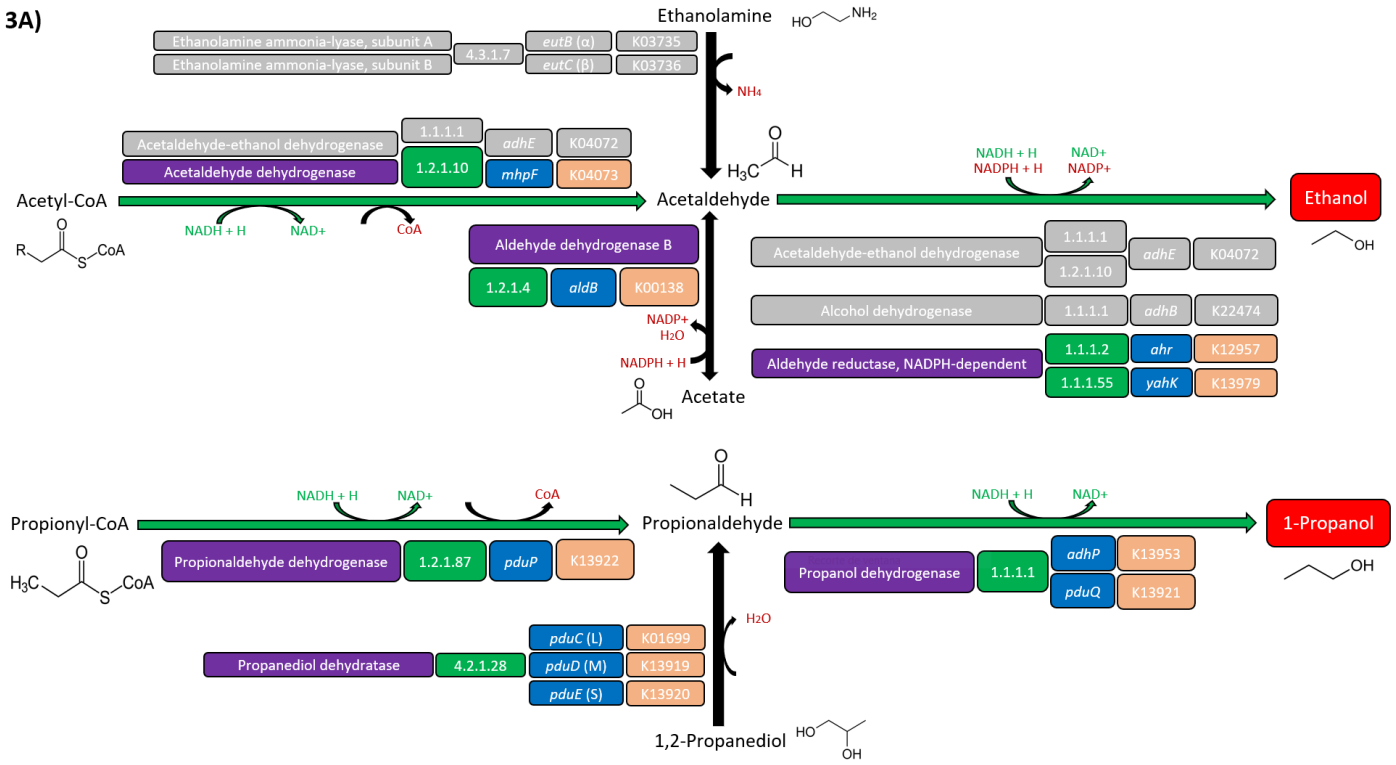
179 **Bacterial genes involved in the formation of SCAs are increased in MASLD**

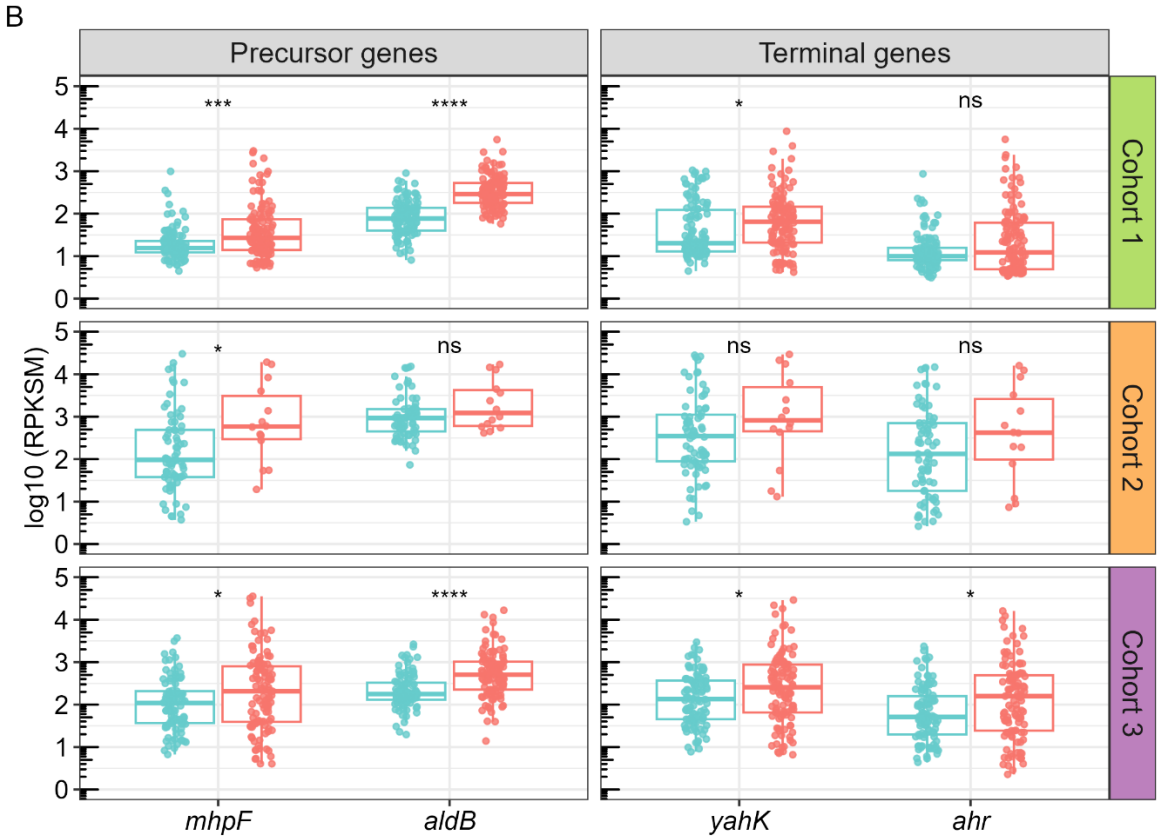
180 Short chain alcohols (SCAs) are a metabolic family also suggested to regulate GLA integrity, with
181 several studies linking GM endogenous alcohol production to MASLD. The first, published over a decade
182 ago, reported elevated peripheral blood ethanol and fecal *Escherichia* spp. 16S-rRNA levels in NASH
183 patients [51]. Similar findings were reported in a pediatric cohort, where children with MASLD exhibited
184 higher levels of fecal ethanol, *Prevotella* spp., and *Gammaproteobacteria* 16S [52]. A more recent study
185 showed that postprandial peripheral blood ethanol and fecal *Lactobacillaceae* 16S-rRNA levels
186 increased in NASH patients after selective inhibition of host ADH, an effect that was reversed with
187 broad-spectrum antibiotic treatment [53]. In 2019, Yuan and colleagues [54] identified a high-alcohol-
188 producing strain of *Klebsiella pneumoniae* capable of causing MASLD through ethanol production.
189 Specifically, they found that FMT from a NASH patient containing this strain induced MASLD in healthy
190 mice, whereas removing the strain before FMT prevented the disease. Furthermore, ethanol produced
191 by this *K. pneumoniae* strain was shown to disrupt fat homeostasis in hepatic cells and cause
192 mitochondrial dysfunction oxidative stress and ROS accumulation [55], which are critical to NASH
193 development as they drive lipid peroxidation and inflammation [56]. Additionally, a recent
194 computational study identified a positive correlation between the abundance of two fermentative
195 pathways in certain *Clostridia* families and fatty-liver disease in patients. These pathways include the
196 production of ethanol from pyruvate and the heterolactic fermentative pathway [57].

197 With this ample evidence, we hypothesized that genes encoding the enzymes involved in the
198 formation of endogenous SCAs could be another metabolic mechanism involved in MASLD pathology.
199 We focused on ethanol and propanol, which are produced by the reduction of the carbonyl group in
200 their corresponding aldehydes. Bacterial ethanol is formed from acetaldehyde through a fermentative
201 reaction that can be catalyzed by an NADH-dependent, bifunctional acetaldehyde-ethanol
202 dehydrogenase encoded by *adhE*, an NADPH-dependent alcohol dehydrogenase encoded by *adhB* [58,
203 59] and a NADPH-dependent aldehyde reductase encoded by two different genes: *ahr* and *yahK*, with
204 the corresponding liberation of phosphorylated reducing power in form of $NADP^+$ [60, 61]. Precursor
205 acetaldehyde can be formed from 1) ethanolamine degradation, through a reaction catalyzed by
206 ethanolamine ammonia-lyase which is encoded by *eutB* and *eutC* genes and liberates ammonia as a by-
207 product [62], 2) acetyl-CoA, through a fermentative reaction catalyzed by bifunctional AdhE and by a
208 CoA-acetylating acetaldehyde dehydrogenase encoded by *mhpF* [63, 64], and 3) acetate, through a
209 dehydration catalyzed by aldehyde dehydrogenase B, encoded by *aldB* [65] (Fig. 3A, top). Bacterial 1-
210 propanol is formed from propionaldehyde through a fermentative reaction catalyzed by a propanol
211 dehydrogenase encoded by two different genes: *adhP* (a propanol-preferring alcohol dehydrogenase)
212 and *pduQ.* Propionaldehyde can be formed by two different reactions: a CoA transference from
213 propionyl-CoA catalyzed by a propionaldehyde dehydrogenase encoded by *pduP*; and a dehydration of
214 1,2 propanediol catalyzed by a propanediol dehydratase composed by three subunits encoded by *pduC*,
215 *pduD* and *pduE* [66] (Fig. 3A, bottom).

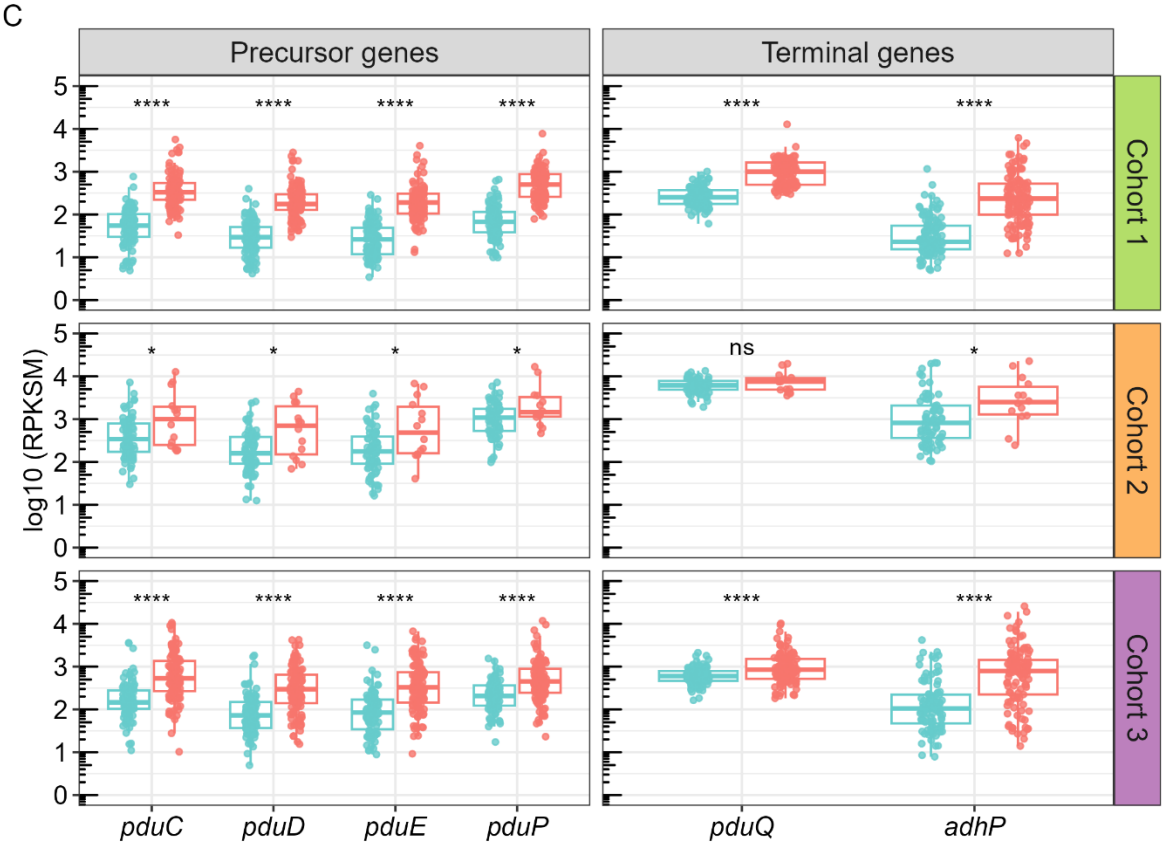
216 We quantified the abundance of genes involved in bacterial SCA formation in the metagenomic
217 samples of the three cohorts as described above. *mhpF* and *aldB*, involved in ethanol production, and
218 *yahK* and *ahr*, involved in acetaldehyde production, were significantly enriched in the GM of MASLD

patients from Cohort 1 (Fig. 3B, up) and Cohort 3 (Fig. 3B, bottom) compared to healthy groups. Their abundance also increased with MASLD progression (Fig. 3B, middle). *mhpF* and *ahr* were depleted compared to *aldB* and *yahK* respectively in all three cohorts. All genes involved in the formation of 1-propanol and propionaldehyde were also significantly enriched in MASLD patients by more than half order of magnitude, increasing with MASLD progression (Fig. 3C). These findings reveal a systematic increase of SCA-producing genes in MASLD.




B
Precursor genes
Terminal genes
Cohort 1
Cohort 2
Cohort 3
log10 (RPKSM)
***
****
*
ns
mhpF
aldB
yahK
ahr

226


C
Precursor genes
Terminal genes
Cohort 1
Cohort 2
Cohort 3
log10 (RPKSM)
****
*
ns
pduC
pduD
pduE
pduP
pduQ
adhP

227

228 **Figure 3. SCA-producing genes are increased in MASLD. (A)** Characterization of the terminal reactions involved in the
229 formation of bacterial ethanol (up) and propanol (down). Enzymes, EC numbers, coding genes and KOs, as well as the
230 reaction type, are indicated for each metabolic step according to the annexed legend in Fig. 2A. Non-candidate genes are
231 colored in grey. **(B)** Abundance of *mhpF, aldB, yahK* and *ahr* genes involved in the last two reactions from pyruvate to ethanol
232 in the three cohorts. **(C)** Abundance of *pduC-E, pduP, pduQ* and *adhP* genes, involved in the last two reactions from 1,2-
233 propanediol and propionyl-CoA to 1-propanol in the three cohorts. Boxplot organization and distribution in facets, as well
234 as statistical tests, were elaborated analogously as Fig. 2C.

235

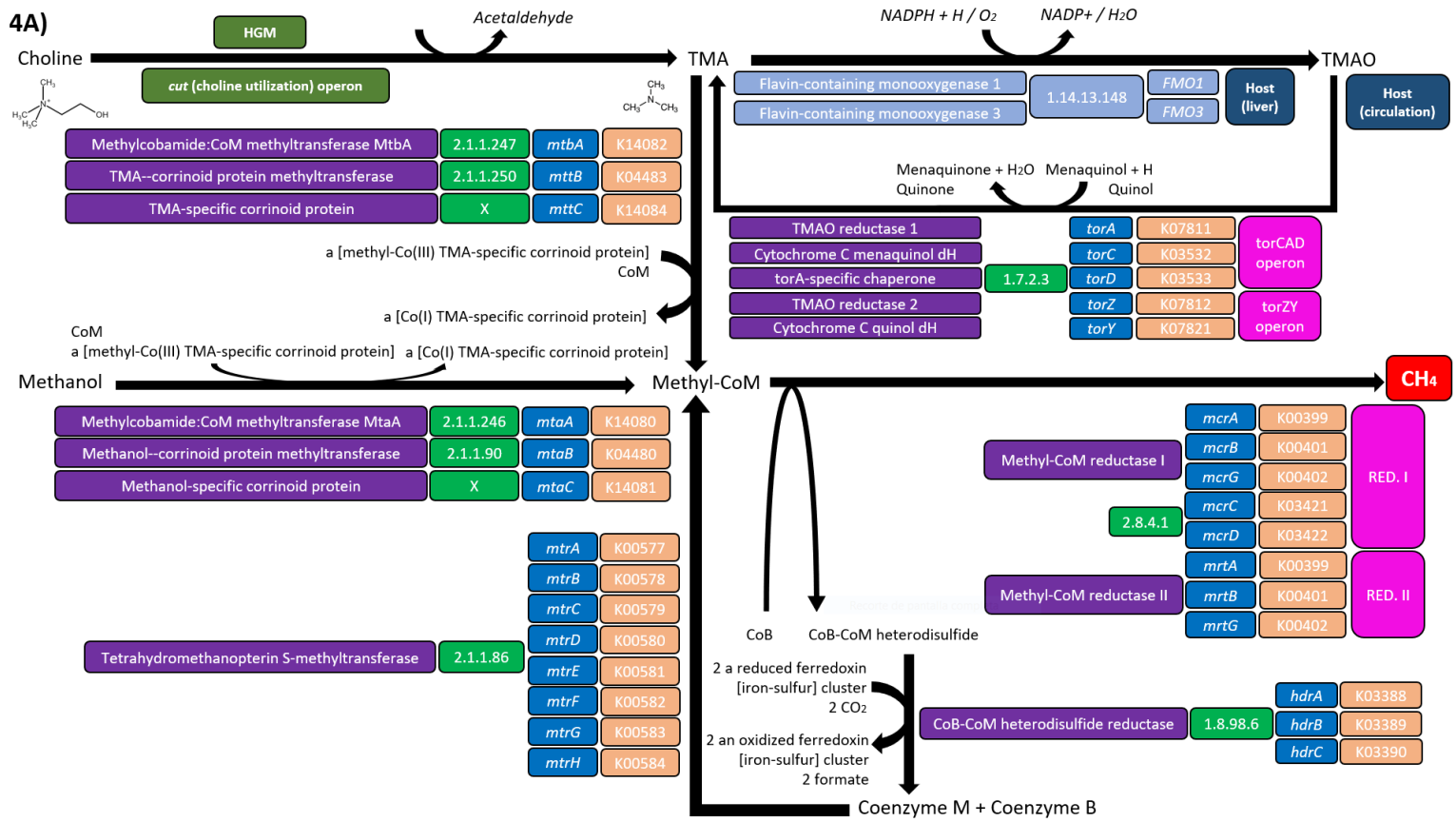
236 **Genes associated with methane production are decreased in MASLD, while *tor* operons driving TMA**
237 **accumulation are elevated**

238 Circulating trimethylamine N-oxide (TMAO) levels have also been linked to MASLD severity [67]
239 and all-cause mortality in MASLD patients [68]. Higher plasmatic TMAO levels also increase
240 cardiovascular disease risk -a condition associated with MASLD- in both mice [69] and clinical patients
241 [70]. Therefore, we investigated whether bacterial TMAO could be another metabolic mechanism
242 involved in MASLD. Methane production in the GM occurs via a two-step process: trimethylamine
243 (TMA) demethylation and CoM methylation, followed by the reduction of methyl-CoM (Fig. 4A). TMA
244 originates from choline degradation through the *cut* operon [71] and can either be oxidized in the liver
245 to TMAO [72] or reduced back to TMA under anaerobic conditions via the *torCAD* and *torZY* operons.
246 Specifically, *torA* and *torZ* encode two TMAO reductases, whose molecular mechanisms are reviewed
247 in [73]. CoM methylation occurs through two well-characterized enzymatic systems: a TMA-specific
248 (composed of *mttC*, *mtbA* and *mttB*) and a methanol-specific pathway (composed of *mtaC, mtaA* and
249 *mtaB*) [74–76]. The final reduction of methyl-CoM to methane is catalyzed by methyl-CoM reductase
250 (MCR), primarily encoded by the *mcrBDCGA* operon [77], with an alternative isozyme (*mrtBGA*) [78].
251 The resulting heterodisulfide (CoB-S-S-CoM) is then reduced to CoM and CoB by a heterodisulfide
252 reductase encoded by operon *hdrABC*, which regenerate both cofactors [79]. Additionally, methyl-CoM

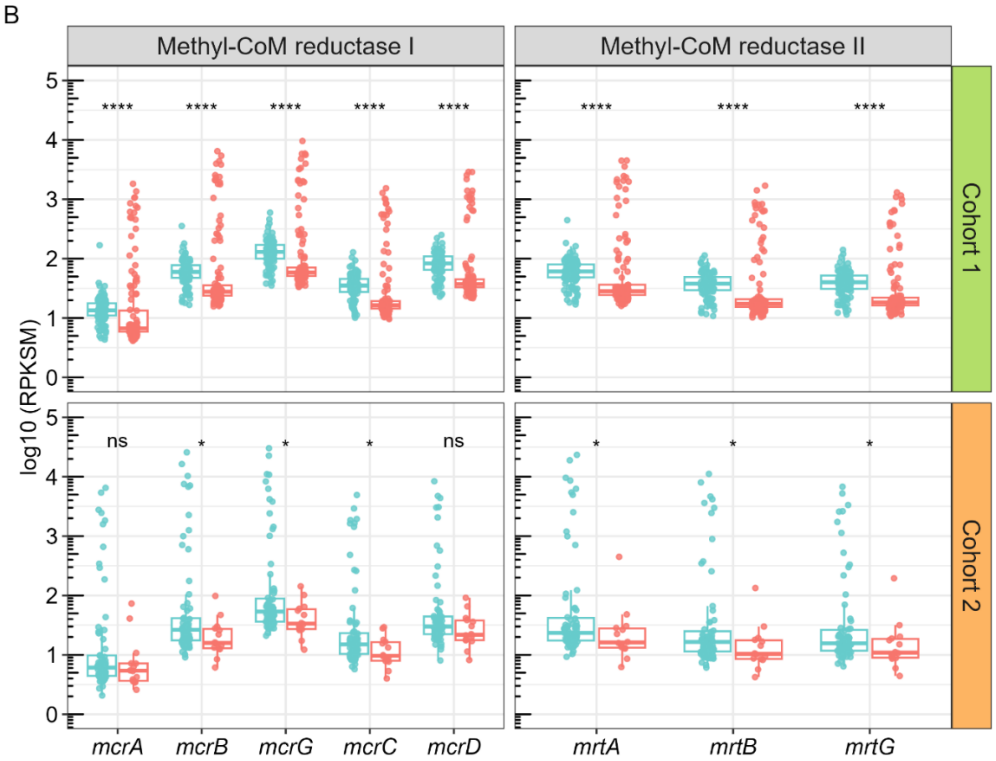
253 is replenished through tetrahydromethanopterin CoM-methyltransferase (encoded by operon
254 *mtrABCDEFGH*), which methylates CoM [80].

255 We measured the abundance of these genes in the samples of the three MASLD cohorts as
256 described above. The MCR I and II coding genes, involved in reducing methyl-CoM to methane,
257 significantly decreased in the GM of MASLD patients (Fig. 4B, up). Their abundance also declined with
258 MASLD progression (Fig. 4B, bottom). Concordantly, *hdr* and *mtr* genes, responsible for replenishing
259 CoM and transferring it to a methyl group to form precursor methyl-CoM, were significantly decreased
260 in MASLD (Fig. 4C, up), and showed a further decrease with MASLD progression (Fig. 4C, bottom).
261 Interestingly perhaps, *hdrA* was two orders of magnitude more abundant than the rest of the genes
262 from the operon. *mtbA, mttB, mtaA* and *mtaB* were significantly decreased in MASLD (Fig. 4E, up).
263 However, they did not exhibit differences between the initial and advanced stages of MASLD (Fig. 4E,
264 bottom). Notably, the higher abundance of methane-producing genes observed in the healthy group
265 was not replicated in Cohort 3. In contrast, bacterial genes involved in TMA regeneration showed an
266 opposite trend. The *torCAD* and *torZY* operons were significantly increased in the GM of MASLD
267 patients, and their abundance increased with MASLD progression (Fig. 4D). These findings suggest that
268 methane production from TMA and methanol that naturally occurs in a healthy state is disrupted in
269 MASLD, revealing another potential microbial pathway involved in MASLD pathogenesis.

270 Finally, we measured the abundance of five housekeeping genes (HKGs), as described in
271 Material and Methods, to confirm that observed differences in genic abundance between sample
272 groups were truly due to MASLD phenotype and not influenced by other biological factors nor spurious
273 technical artifacts derived from the quantification method. We found no differences in gene abundance
274 between groups for any HKG in any of the cohorts (Supplementary Fig. 1).

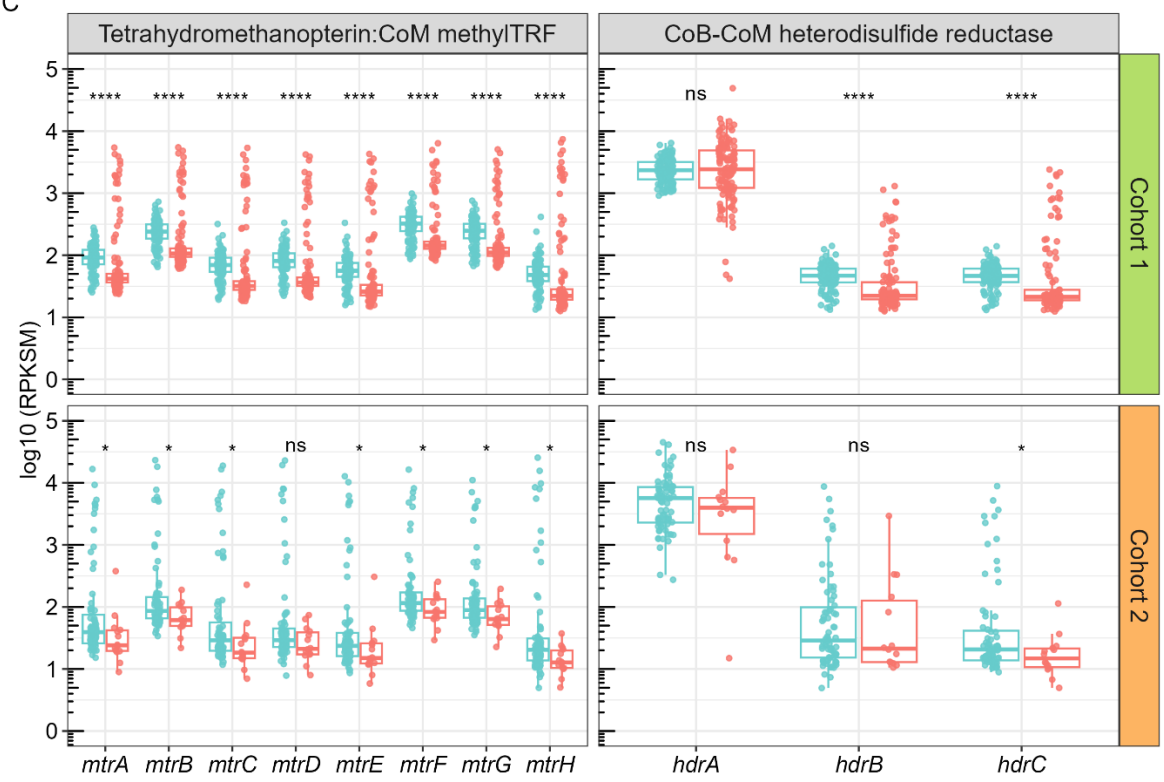

4A)
HGM
Acetaldehyde
Choline
TMA
NADPH + H / O2
NADP+ / H2O
TMAO
cut (choline utilization) operon
Flavin-containing monooxygenase 1
Flavin-containing monooxygenase 3
1.14.13.148
FMO1
FMO3
Host (liver)
Host (circulation)
Methylcobamide:CoM methyltransferase MtbA
2.1.1.247
mtbA
K14082
TMA--corrinoid protein methyltransferase
2.1.1.250
mttB
K04483
TMA-specific corrinoid protein
X
mttC
K14084
Menaquinone + H2O
Quinone
Menaquinol + H
Quinol
TMAO reductase 1
Cytochrome C menaquinol dH
torA-specific chaperone
TMAO reductase 2
Cytochrome C quinol dH
1.7.2.3
torA K07811
torC K03532
torD K03533
torZ K07812
torY K07821
torCAD operon
torZY operon
a [methyl-Co(III) TMA-specific corrinoid protein]
CoM
a [Co(I) TMA-specific corrinoid protein]
CoM
a [methyl-Co(III) TMA-specific corrinoid protein]
a [Co(I) TMA-specific corrinoid protein]
Methanol
Methyl-CoM
CH4
Methylcobamide:CoM methyltransferase MtaA
2.1.1.246
mtaA
K14080
Methanol--corrinoid protein methyltransferase
2.1.1.90
mtaB
K04480
Methanol-specific corrinoid protein
X
mtaC
K14081
Methyl-CoM reductase I
mcrA K00399
mcrB K00401
mcrG K00402
mcrC K03421
mcrD K03422
2.8.4.1
RED. I
Methyl-CoM reductase II
mrtA K00399
mrtB K00401
mrtG K00402
RED. II
mtrA K00577
mtrB K00578
mtrC K00579
mtrD K00580
mtrE K00581
mtrF K00582
mtrG K00583
mtrH K00584
Tetrahydromethanopterin S-methyltransferase
2.1.1.86
CoB
CoB-CoM heterodisulfide
2 a reduced ferredoxin [iron-sulfur] cluster
2 CO2
2 an oxidized ferredoxin [iron-sulfur] cluster
2 formate
CoB-CoM heterodisulfide reductase
1.8.98.6
hdrA K03388
hdrB K03389
hdrC K03390
Coenzyme M + Coenzyme B

275


B
Methyl-CoM reductase I
Methyl-CoM reductase II
Cohort 1
Cohort 2
log10 (RPKSM)
****
ns
*
mcrA
mcrB
mcrG
mcrC
mcrD
mrtA
mrtB
mrtG

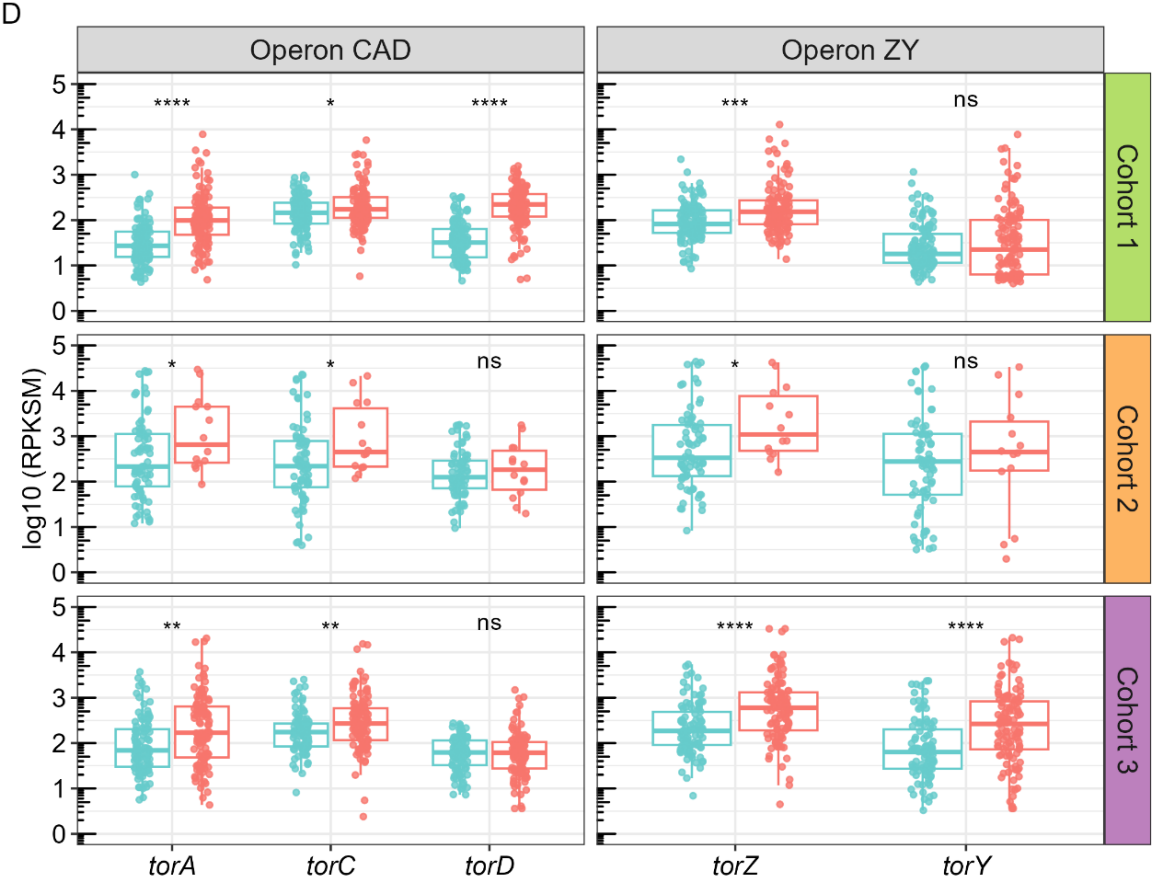
276

C


Tetrahydromethanopterin:CoM methylTRF
CoB-CoM heterodisulfide reductase
Cohort 1
Cohort 2
log10 (RPKSM)
mtrA
mtrB
mtrC
mtrD
mtrE
mtrF
mtrG
mtrH
hdrA
hdrB
hdrC
ns

277

D


Operon CAD
Operon ZY
Cohort 1
Cohort 2
Cohort 3
log10 (RPKSM)
torA
torC
torD
torZ
torY
ns

278

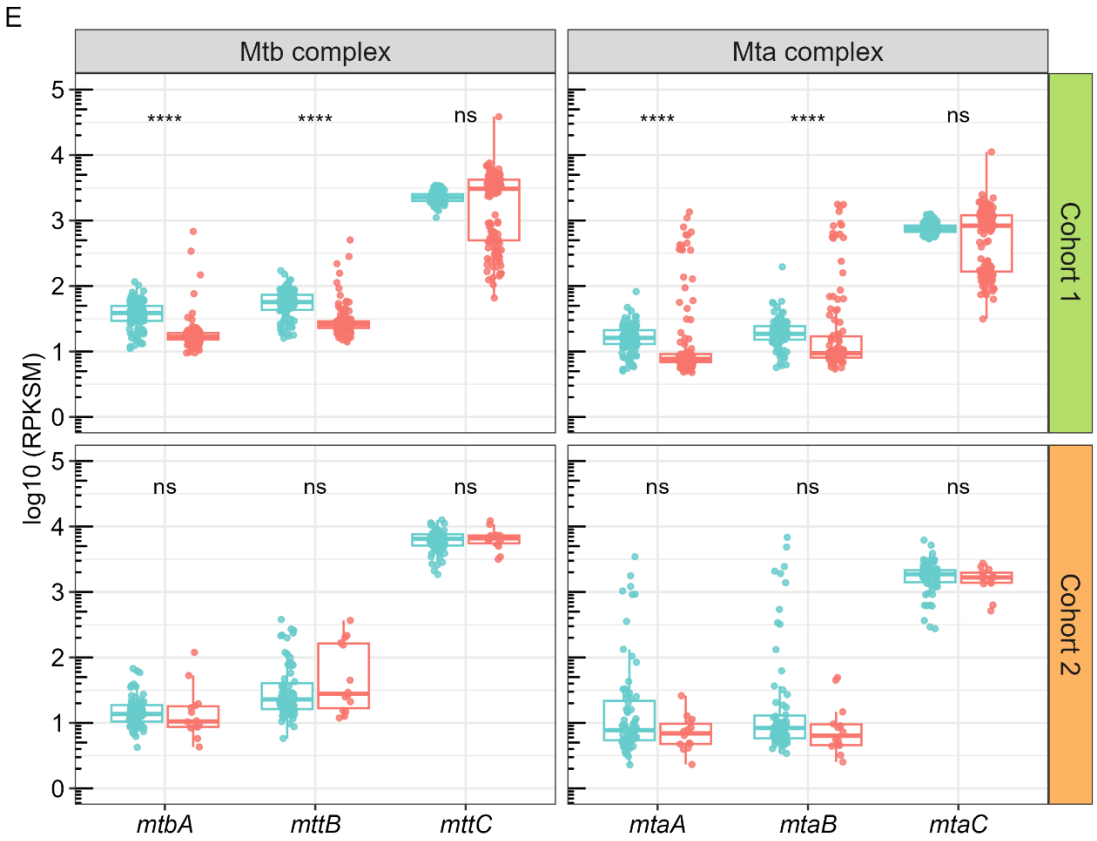


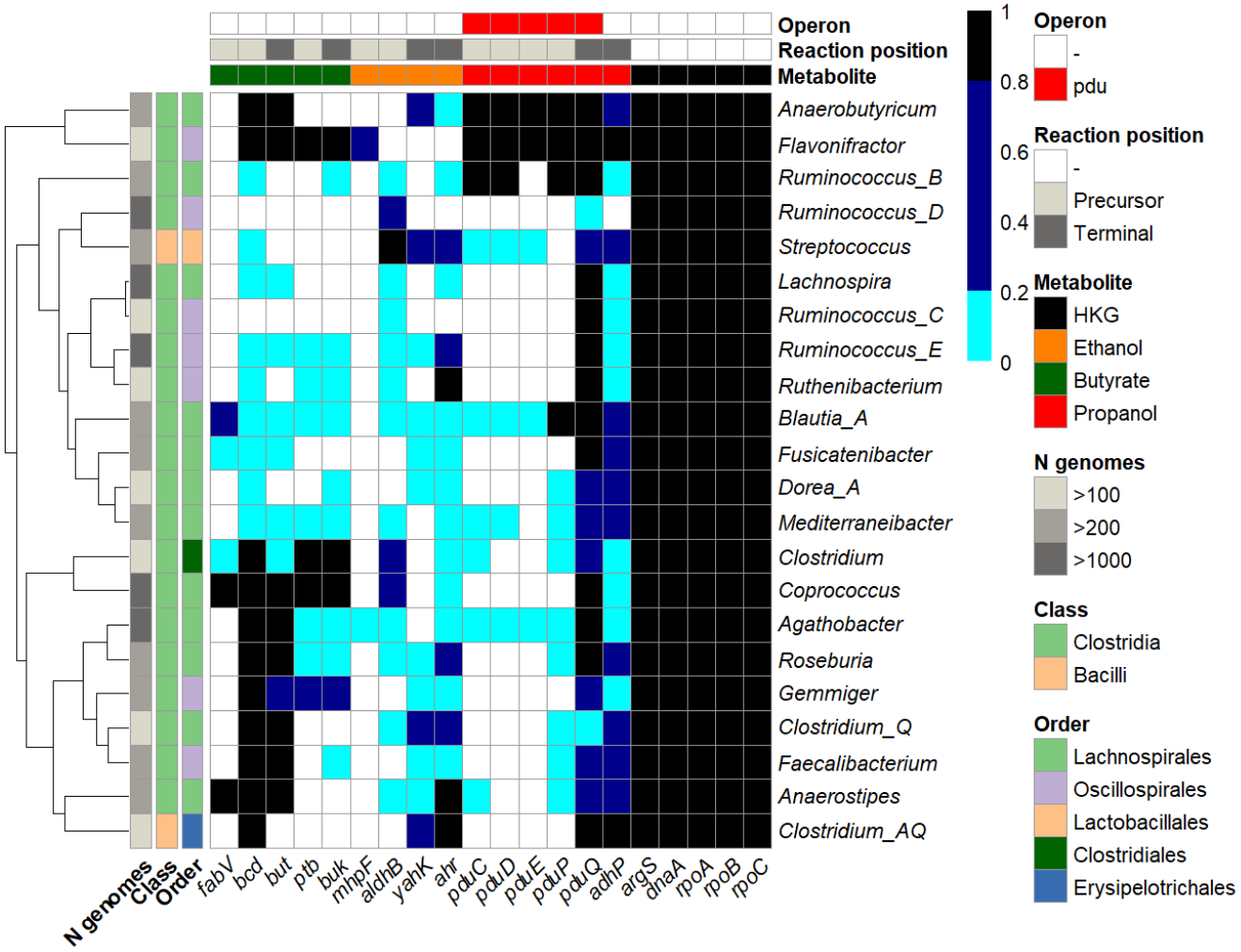
Figure 4. Genes involved in methane production are decreased in MASLD, whereas *tor* operons responsible for TMA accumulation are increased in MASLD. (A) Characterization of the terminal reactions involved in the formation of microbial methane. Enzymes, EC numbers, coding genes and KOs, as well as the reaction type, are indicated for each metabolic step according to the annexed legend in Fig. 2A. **(B)** Abundance of genes coding for methyl-coM reductase I (*mcr*) and methyl-CoM reductase II (*mrt*). **(C)** Abundance of genes coding for CoB-CoM heterodisulfide reductase (*hdr*) and tetrahydromethanopterin S-methyltransferase (*mtr*). **(D)** Abundance of *tor* operons. **(E)** Abundance of genes coding for the Mtb and Mta complexes. Facets represent genic operons. Boxplot organization and statistical tests were elaborated analogously as Fig. 2C.

Candidate metabolic genes are accessory in the GM

Another possible explanation for the absence of a consistent taxonomic signature associated with MASLD is that the 16S rRNA gene is probably not a sufficiently robust marker to unravel a causal link between GM composition and MASLD phenotype through metataxonomic studies. GM bacterial species have pleomorphic genomes, consisting of up to thousands of genomically distinct strains that share a conserved genetic core but possess an accessory genome that is highly variable among them

295 [81, 82]. This accessory genome is often encoded in plasmids and other mobile genetic elements which
296 are subject to frequent changes and can readily transfer between different bacterial strains [15, 83].
297 To test this hypothesis, we assessed whether candidate metabolic genes involved in MASLD
298 were core or accessory across human GM genomes. Their presence was determined as described in
299 Materials and Methods. Briefly, we defined the presence of a target gene as the percentage of strains
300 within a given genus that encode it. A gene was classified as core if present in >80% of congeneric
301 genomes, as accessory if present in 20-80%, and as highly accessory if present in <20% genomes. The
302 most abundant genera in the GM (n=22) belong to class Clostridia, specifically to orders *Lachnospirales*
303 and *Oscillospirales* (Fig. 6A). Butyrate-producing genes from the butyryl-CoA pathway (*bcd* and *but*)
304 were core in nearly half of these genera, while those from the butyryl-P pathway (*ptb* and *buk*) were
305 core only in *Flavonifractor, Clostridium* and *Coprococcus.* Notably, butyrate-producing genes were
306 highly accessory across several genera, including *Lachnospira, Ruminococcus_B* and *E,*
307 *Ruthenibacterium, Blautia_A*, *Fusicatenibacter*, *Dorea_A,* and *Mediterraneibacter*, and were
308 completely absent in *Ruminococcus_C* and *D*. In contrast, acetaldehyde and ethanol-producing genes
309 were highly accessory across all genera except *Streptococcus*, an heterolactic genus that constitutively
310 encoded *aldhB* in all its genomes (specifically in *S. epidermis, S. salivarius* and *S. thermophilus,*
311 Supplementary Fig. 2) and exhibited moderated presence (20-80%) of *yahK*/*ahr*. The acetaldehyde-
312 producing gene *mhpF* was nearly absent in the GM, except for a few genomes of *Flavonifractor* and
313 *Agathobacter*. Propanaldehyde and propanol-producing genes from the *pdu* operon were core in three
314 genera: *Ruminococcus_B* (*R_B. gnavus*), *Anaerobutyricum* (*A. hallii*) and *Flavonifractor* (*F. plautii*),
315 though *pduE* was unexpectedly absent in *Ruminococcus_B*. These genes were highly accessory in
316 *Blautia_A* (*B_A. wexlerae*), *Streptococcus*, *Medierraneibacter* or *Agathobacter* (*A. rectalis*, formerly

317 *Eubacterium rectale*). Notably, *A. hallii* and *F. plautii* were the only species encoding both core butyrate
318 and propanol-producing genes. See Supplementary Fig. 2 for within-species distribution.
319 Beyond their accessory nature within GM pangenomes, candidate metabolic genes were also
320 encoded in plasmids (Fig. 6B). Genes involved in ethanol formation such as *aldhB, yahK* and *ahr* were
321 among the most widely distributed, being present in hundreds of different plasmids. In contrast,
322 butyrate-producing genes (*fabV, but, ptb,* and *buk*) and propanol-related genes (*pduC, pduD, pduE,*
323 *pduP,* and *pduQ*) were detected in fewer plasmids, further supporting their variable presence across
324 GM genomes. A notable exception was *bcd*, which exhibited a broader presence across both
325 chromosomal and plasmid-encoded sequences. These findings reinforce the idea that non-essential
326 metabolic functions, especially ethanol production, are encoded in highly mobile genetic elements,
327 contributing to their accessory nature within the GM.
328



329

Gene	HMM size (aa)	N gene (+) plasmids
fabV	394	46
bcd	379	819
but	446	36
ptb	303	120
buk	358	3
mhpF	312	144
aldhB	512	1,203
yahK	350	871
ahr	339	538
pduC	553	31
pduD	224	2
pduE	174	2
pduP	461	6
pduQ	370	290
adhP	336	1,007

330

Figure 5. Candidate metabolic genes involved in MASLD are predominantly accessory in the human GM. (A) Heatmap showing the presence of candidate genes involved in MASLD across the most abundant genera in the human GM. The x-axis represents genes associated with butyrate and SCAs production, as well as HKGs. The y-axis lists the most abundant genera. Genic presence is expressed as the percentage of congeneric genomes encoding each target gene, with colors indicating classification: black for core genes (>80% of genomes), dark blue for accessory genes (20-80% genomes), light blue for highly accessory genes (<20% genomes) and white for genes completely absent in the genus. Hierarchical clustering was applied to the y-axis to group genera based on genic presence similarity. **(B)** Presence of candidate genes in RefSeq200 annotated plasmids. Column 3 represents the number of different plasmids that encode at least one copy of each gene.

3. Discussion

MASLD is a liver pathology with multiple factors involved. One of them is GM and its associated metabolism, which causes an effect on GLA integrity. Several studies have tried to ascribe a causality between GM composition and MASLD. However, there are many inconsistent taxonomic signatures in key GM clades such as *Blautia spp.* or *Roseburia spp.* whose abundance has been positively and negatively correlated to MASLD progression across multiple studies [27]. To confirm if there is a specific clade associated with MASLD, we profiled the taxonomic composition of the GM present in the fecal samples of healthy and MASLD donors from three different cohorts of study (Fig. 1). Although we identified five candidate bacterial genera showing consistent differences in their relative abundance between groups across all the cohorts, only an unclassified *Lachnospiraceae* showed statistically significant differences between groups (Fig. 1C). This result indicates that there is no clear-cut relationship between alterations in GM composition and MASLD.

16S rRNA profiling in metataxonomic studies provides a limited view of GM changes [27]. Due to bacterial genome plasticity, the GM exhibits vast strain diversity [81] which remains undetected by 16S rRNA abundances despite its potential impact on metabolic functions. Previously, we reported that geno-metabolic shifts in GM propionate production associated with inflammatory bowel disease did not align with species-level variations [84]. Given the role of metabolic profiles in phenotypic

357 differences, we focused on identifying abundance differences in the genes responsible for producing
358 metabolites potentially implicated in MASLD across several patient cohorts.

359 Butyrate-producing genes are depleted in the GM of MASLD patients (Fig. 2B), particularly those
360 involved in the crotonyl-butyryl CoA axis: *fabV* (butyryl-CoA oxidoreductase), *bcd* (butyryl-CoA
361 dehydrogenase) and *but* (butyryl-CoA:acetate-CoA transferase, Fig. 2A). While *ptb* and *buk* where also
362 higher in both healthy subcohorts, their levels did not differ significantly between initial and advanced
363 MASLD groups within Cohort 2. This suggests that CoA transfer is the predominant pathway for butyrate
364 production in the GM, as previously reported [85]. However, the smaller differences in *ptb* and *buk* in
365 Cohort 2 may result from a shared MASLD phenotype across different fibrosis stages, coupled with a
366 smaller sample size compared to Cohort 3. Additionally, two other enzymes -structured as two-subunit
367 systems-, can transfer CoA from different cofactors to butyrate to produce butyryl-CoA: butyryl-
368 CoA:acetoacetate-CoA transferase (encoded by *atoD* and *atoA*) and acetate CoA-transferase (encoded
369 by by *ctfA* and *ctfB*). These were not included in this study as they catalyze the reverse reaction,
370 consuming butyrate to form butyryl-CoA [86]. Notably, the depletion of butyrate genes was not
371 observed in MASLD patients from Cohort 1.

372 Genes involved in the formation of ethanol and propanol through connected fermentative
373 reactions (Fig. 3A) are increased in MASLD. Genes *mhpF* and *aldB* -encoding two aldehyde
374 dehydrogenases that produce ethanol-, and *yahK* and *ahr* -encoding two NADPH-dependent aldehyde
375 reductases that form acetaldehyde- are significantly increased in the GM of MASLD patients compared
376 to the healthy groups (Fig. 3B). Genes involved in the production of 1-propanol structured in operon
377 *pdu* are also increased in MASLD (Fig. 3C). Specifically, *pduCDE* -encoding a propanediol dehydratase-,

378 *pduP* -encoding a propionaldehyde dehydrogenase- and *pduQ* and *adhP* -encoding two distinct
379 propanol dehydrogenases-. The marked differential abundances of propanol-producing genes,
380 organized in the *pdu* operon and *adhP* and unexplored so far, suggest that propanol could play an even
381 more critical role than ethanol in MASLD onset and progression.

382 The abundance of genes in the TMA-methane metabolic axis is disrupted in MASLD. Genes
383 involved in methane formation (Fig. 4A) are depleted in the pathology, including both MCR I and II
384 coding-genes (Fig. 4B), which reduce methyl-CoM to methane, and *mtr* and *hdr* genes (Fig. 4C),
385 dedicated to regenerating methyl-CoM. In contrast, TMAO-reductive genes, organized in *torCAD* and
386 *torZY* operons and responsible for regenerating TMA, are increased in MASLD (Fig. 4D). These results
387 suggest that methane production is impaired in MASLD. The opposing trend in the abundance of *tor*
388 operons further indicates that this impairment stems from a metabolic bottleneck that leads to TMA
389 accumulation instead of its conversion to methyl-CoM, effect that is driven by TMAO reductases I and
390 II.

391 Finally, we determined that metabolic genes associated with MASLD exhibit a highly accessory
392 profile within the GM (Fig. 5A). Butyrate-producing genes, particularly those from the butyryl-CoA
393 pathway (*bcd* and *but*), are highly accessory in nearly half of the most abundant GM genera, while SCA-
394 producing genes display an even more pronounced accessory profile. In fact, butyryl-CoA pathway
395 genes are more pervasive in the GM than those from the butyryl-P pathway. The previously reported
396 absence of a consistent taxonomic signature in MASLD [27], also observed in this study (Fig. 1) likely
397 stems from this genomic plasticity. Strain-level metabolic variability in the GM generates substantial
398 phenotypic differences within species, leading to variable virulence or metabolic profiles, as previously

399 observed in pathogenic and commensal *E. coli* strains [87, 88] Further supporting their mobility,
400 candidate genes are frequently encoded in plasmids (Fig. 5B) with ethanol formation genes (*aldhB*,
401 *yahK*, and *ahr*) being the most widely distributed, while butyrate and propanol-related genes are
402 present to a lesser extent. Since non-essential, adaptive metabolic pathways are often encoded in
403 plasmids across many bacterial species [88–90], their presence likely drives phenotypic differences in
404 MASLD without altering the overall species composition detected by conventional 16S rRNA-based
405 metataxonomic studies. Given that MASLD is partially driven by bacterial metabolism, strains from
406 different species probably may produce similar liver-impacting metabolites, potentially contributing to
407 MASLD pathogenesis.

408

409 **4. Conclusion**

410 Our study demonstrates that the gut microbiome's contribution to MASLD extends beyond
411 taxonomic composition and is deeply rooted in strain-level functional diversity shaped by genome
412 plasticity and horizontal gene transfer. While species like *Agathobacter rectalis* emerge as consistent
413 taxonomic indicators of disease, it is the accessory, plasmid-encoded metabolic genes -particularly
414 those involved in the production of butyrate, ethanol, and propanol- that better capture the metabolic
415 reprogramming characteristic of MASLD. These findings underscore the limitations of traditional
416 marker-based taxonomic profiling and highlight the need to integrate functional metagenomics and
417 mobile genetic element analyses to fully understand microbiome-disease interactions. By identifying
418 metabolically relevant microbial functions consistently altered in MASLD, our work provides new

419 insights into disease mechanisms and points toward strain-level and pathway-specific targets for
420 diagnostic and therapeutic development.

421

422 **Materials and Methods**

423 ***Downloading of shotgun-sequencing metagenomic data***

424 A total of 554 whole shotgun metagenomic libraries generated from the fecal samples of MASLD
425 and healthy subjects were downloaded from the ENA using the accession numbers corresponding to
426 five different studies, and they were stratified in three MASLD cohorts as detailed in Table 1.
427 Information on the collection and processing of fecal samples, bacterial DNA extraction and sequencing
428 protocols can be found in the original studies. Additionally, we retrieved the medical metadata
429 associated with each sample donor to confirm their MASLD status and classify samples into two
430 phenotypic groups within each cohort. Samples lacking a clearly defined MASLD status were excluded
431 from this study. All data was downloaded between April and March 2023.

432

433 ***Description of MASLD patient cohorts***

434 Cohort 1 consists of two sets of European, obese MASLD patients aged 20-64 years with initial
435 fibrosis (F0-F2) and varying degrees of steatosis: 73 patients from [18] and 10 patients from [19]. In the
436 latter group, 48 fecal samples were collected at five time points during the study (two samples from
437 day 1 were not obtained). Diagnoses were confirmed through hepatic biopsy performed during obesity
438 surgery, and none of the subjects had alcoholic liver disease, hepatitis or diabetes. The control group
439 for comparison includes 117 healthy patients from [17]. Cohort 2 comprises 86 American MASLD

440 patients from [20], including 72 with initial (F0-F2) and 14 with advanced fibrosis (F3-F4). Diagnoses
441 were confirmed through hepatic biopsy and magnetic resonance imaging, and none of the patients had
442 alcoholic liver disease, hepatitis or diabetes. In this manuscript, the initial fibrosis group is referred to
443 as "healthy" for group comparison purposes, despite including patients with early-stage MASLD. Cohort
444 3 includes 230 Han Chinese patients from [21]: 125 diagnosed with cirrhotic MASLD and 105 with no
445 hepatic pathology. None of these subjects had hepatitis. All cohorts consist of one metagenomic stool
446 sample per patient, except for the 48 samples collected at multiple time points from 10 patients in the
447 [19] study.
448
449 ***Quality control of metagenomic reads***
450 Shotgun sequencing reads were trimmed to remove Illumina adapter remnants and low-quality
451 regions (<Q25) using BBDuk from the BBTools suite (version 37.62)
452 (https://sourceforge.net/projects/bbmap/) [91]. Read quality was assessed with FastQC (version
453 0.12.1) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) [92] and summarized using
454 MultiQC (version 1.22.3) [93].
455
456 ***Taxonomic signatures associated with MASLD***
457 To determine whether any bacterial clade is systematically increased or decreased in MASLD
458 across the three cohorts, we profiled the taxonomic composition of GM communities in fecal samples
459 using MetaPhlAn (version 4.0.6) [28, 29]. Briefly, MetaPhlAn estimates the relative abundance of
460 bacterial genera by aligning the metagenomic reads against a curated database of 5 million unique
461 clade-specific marker genes derived from 1 million microbial genomes spanning 26,970 species-level

462 genome bins. For this study, we used the database version *mpa_vOct22_CHOCOPhlAnSGB_202212*,
463 downloaded on August 22, 2023, and added the *--ignore_eukaryotes* flag to skip profiling eukaryotic
464 organisms. Post-processing of the MetaPhlAn output involved removing low-abundant clades (i.e.,
465 those with an average relative abundance of 0), excluding genera predicted as "unclassified", and
466 verifying that excluded genera were not exclusively present in only one of the comparison groups (Fig.
467 1).
468

469 ***Isolation of target gene families in the GM***

470 We constructed families of the human GM-metabolic genes encoding the enzymes involved in the
471 production of butyrate, ethanol, propanol, TMA, methane and their precursor metabolites, as
472 represented in Figs. 2-4. The Unified Human Gastrointestinal Genome (UHGG, version 2.0.2) [81] served
473 as the reference database for this purpose. This resource, comprising over 289,000 bacterial and
474 archaeal genomes and more than 665 million proteins, represents the most comprehensive collection
475 of characterized human GM genes to date. Each gene family was built by retrieving a seed of sequences
476 based on gene and protein annotation from the UHGG, assessing their similarity with lab-validated
477 enzymatic sequences when available. Validated gene sequences were retrieved from MetaCyc [94] and
478 used as controls to assess their phylogenetic proximity to the retrieved gene seeds. Gene families were
479 aligned using MAFFT (version 7.271) [95] and phylogenetic trees were constructed with IQ-TREE
480 (version 2.0.3) [96]. Low phylogenetic distances between control and retrieved sequences served as a
481 criterion for defining gene families. Duplicated sequences, gene fragments and phylogenetically distant
482 genes were excluded after inspecting the trees. To improve gene identification and classification, we
483 then built Hidden Markov Model (HMM) profiles for each target gene family using *HMMER hmpress*

484 (version 3.4) (http://hmmer.org/) [97]. MetaCyc and KEGG [94, 98] were also used to validate gene and
485 enzyme names, KEGG Orthology groups (KOs), the biochemical reactions involved in the formation of
486 each metabolite, and their enzyme commission (EC) numbers. The same protocol was applied to isolate
487 five HKG families, universal single-copy marker genes in prokaryotes, used as controls. These HKGs and
488 their encoded proteins were: *argS* (arginyl-tRNA synthetase), *dnaA* (chromosomal replication initiator
489 protein DnaA), *rpoA*, *rpoB* and *rpoC* (alpha, beta and gamma subunits of the DNA-directed RNA
490 polymerase, respectively).
491

492 ***Analysis of gene abundance by read-based metagenomic quantification***

493 We quantified the abundance of target genes in the metagenomic samples from the three cohorts
494 by aligning the quality-controlled reads against the isolated gene families using DIAMOND (version
495 2.0.14) [99]. Alignments were filtered to retain only the best match for each read and gene family
496 sequence by using the *"--max-hsps 1"* option, with an E-value threshold $<1e^{-10}$. These parameters were
497 based on experimentally validated recommendations for detecting genes in fecal metagenomic
498 samples with DIAMOND [100]. The total number of aligned reads for each gene family was normalized
499 using a modified RPKM formula, referred to as RPKSM. This measure was used as an indicator of gene
500 abundance, and it is defined as the number of aligned reads per kilobase per gene family size per million
501 reads, to account for variations in gene length, gene family size (i.e., the number of sequences in each
502 gene family) and library size, respectively (Figs. 2-4, Supplementary Fig. 1).
503

504

505 ***Presence of candidate metabolic genes across the human GM genomes and annotated plasmids***

506 To determine whether candidate metabolic genes involved in MASLD were core or accessory across
507 the human GM, we analyzed their presence in both UHGG genomes and curated RefSeq200 plasmids
508 [101]. Since many UHGG genomes originate from metagenome-assembled or single-cell-amplified
509 genomes, they may be incomplete, potentially leading to false positive annotations. To minimize this
510 risk, we included only UHGG genomes with >95% completeness and containing the five HKGs described
511 above. ORFs were predicted using Prodigal (version 2.6.3) [102], retaining only those >90% of the HMM
512 size. We then used HMMER *hmmscan* with the following parameters: protein-pHMM coverage
513 threshold >90%, E-value <0.01 and independent E-value (i-E value) <0.01, strictly enforcing the 90%
514 threshold to avoid capturing ORFs with similar protein domains. This approach yielded 31,227 genomes
515 classified as complete and containing the five HKGs.

516 To classify genes as core or accessory, we defined their presence as the percentage of strains within
517 a given genus that encode the gene, considering only the most abundant genera in the GM (Fig. 1) with
518 >100 genomes meeting the completeness criteria. Same process was replicated at species level
519 (Supplementary Fig. 2). Genes were categorized as core if present in >80% of congeneric genomes, as
520 accessory if present in 20-80%, and as highly accessory if present in <20% genomes (Fig. 6A).
521 Additionally, we assessed whether these genes could be encoded in annotated plasmids. Using the
522 same constraints, we predicted ORFs from the RefSeq200 plasmid collection, which includes 23,309
523 plasmids encoding over 2 million ORFs. ORFS were retained if they met the >90% HMM size threshold,
524 with *hmmscan* parameters identical to those applied in the UHGG genome analysis (Fig. 6B).

525

526 ***R packages and statistical analysis***

527 Data manipulation was performed using R 4.1.3 and the *tidyverse* package (version 1.3.2) [103].
528 Figures were generated with the R packages *ggplot2* (version 3.4.2) [104], *ggpubr* (version 0.6.0) [105]
529 and *pheatmap* (version 1.0.12) [106]. Mann-Whitney tests with a Benjamini-Hochberg adjustment for
530 multiple testing were applied on the relative abundance of bacterial genera (Fig. 1C) and genic
531 abundance (Figs. 2B, 3B-C, 4B-E, 5) between phenotypic groups in the metagenomic fecal samples of
532 each cohort to assess the significance of the differences. The statistical analysis was conducted using
533 the R package *rstatix* (version 0.7.2) [107].

534

535 **References**

536 1. Riazi K, Azhari H, Charette JH, et al (2022) The prevalence and incidence of NAFLD worldwide: a systematic review
537 and meta-analysis. Lancet Gastroenterol Hepatol 7:. https://doi.org/10.1016/S2468-1253(22)00165-0

538 2. Marchesini G, Day CP, Dufour JF, et al (2016) EASL-EASD-EASO Clinical Practice Guidelines for the management of
539 non-alcoholic fatty liver disease. J Hepatol 64:1388–1402. https://doi.org/10.1016/j.jhep.2015.11.004

540 3. Allen AM, Therneau TM, Ahmed OT, et al (2022) Clinical course of non-alcoholic fatty liver disease and the
541 implications for clinical trial design. J Hepatol 77:1237–1245. https://doi.org/10.1016/j.jhep.2022.07.004

542 4. Powell EE, Wong VWS, Rinella M (2021) Non-alcoholic fatty liver disease. The Lancet 397:2212–2224

543 5. Yki-Järvinen H (2014) Non-alcoholic fatty liver disease as a cause and a consequence of metabolic syndrome. Lancet
544 Diabetes Endocrinol 2:901–910

545 6. Martín-Mateos R, Albillos A (2021) The Role of the Gut-Liver Axis in Metabolic Dysfunction-Associated Fatty Liver
546 Disease. Front Immunol 12

547 7. Carpino G, Ben M Del, Pastori D, et al (2020) Increased Liver Localization of Lipopolysaccharides in Human and
548 Experimental NAFLD. Hepatology 0: https://doi.org/10.1002/hep.31056/suppinfo

549 8. Li Q, Rempel JD, Yang J, Minuk GY (2022) The Effects of Pathogen-Associated Molecular Patterns on Peripheral Blood
550 Monocytes in Patients with Non-alcoholic Fatty Liver Disease. J Clin Exp Hepatol 12:808–817.
551 https://doi.org/10.1016/j.jceh.2021.11.011

552 9. Miele L, Valenza V, La Torre G, et al (2009) Increased intestinal permeability and tight junction alterations in
553 nonalcoholic fatty liver disease. Hepatology 49:1877–1887. https://doi.org/10.1002/hep.22848

554 10. Wang W, Zhao J, Gui W, et al (2018) Tauroursodeoxycholic acid inhibits intestinal inflammation and barrier
555 disruption in mice with non-alcoholic fatty liver disease. Br J Pharmacol 175:469–484.
556 https://doi.org/10.1111/bph.14095

557 11. Kuang J, Wang J, Li Y, et al (2023) Hyodeoxycholic acid alleviates non-alcoholic fatty liver disease through modulating
558 the gut-liver axis. Cell Metab 35:1752-1766.e8. https://doi.org/10.1016/j.cmet.2023.07.011

559 12. Yang C, Wu J, Yang L, et al (2024) Altered gut microbial profile accompanied by abnormal short chain fatty acid
560 metabolism exacerbates nonalcoholic fatty liver disease progression. Sci Rep 14:22385.
561 https://doi.org/10.1038/s41598-024-72909-8

562 13. Chierico F Del, Nobili V, Vernocchi P, et al (2016) Gut Microbiota Profiling of Pediatric Nonalcoholic Fatty Liver
563 Disease and Obese Patients Unveiled by an Integrated Meta-omics-Based Approach.
564 https://doi.org/10.1002/hep.28572/suppinfo

565 14. Lee G, You HJ, Bajaj JS, et al (2020) Distinct signatures of gut microbiome and metabolites associated with significant
566 fibrosis in non-obese NAFLD. Nat Commun 11:. https://doi.org/10.1038/s41467-020-18754-5

567 15. Redondo-Salvo S, Fernández-López R, Ruiz R, et al (2020) Pathways for horizontal gene transfer in bacteria revealed
568 by a global map of their plasmids. Nat Commun 11:3602. https://doi.org/10.1038/s41467-020-17278-2

569 16. Leinonen R, Akhtar R, Birney E, et al (2011) The European nucleotide archive. Nucleic Acids Res 39:.
570 https://doi.org/10.1093/nar/gkq967

571 17. Poyet M, Groussin M, Gibbons SM, et al (2019) A library of human gut bacterial isolates paired with longitudinal
572 multiomics data enables mechanistic microbiome research. Nat Med 25:1442–1452.
573 https://doi.org/10.1038/s41591-019-0559-3

574 18. Hoyles L, Fernández-Real JM, Federici M, et al (2018) Molecular phenomics and metagenomics of hepatic steatosis
575 in non-diabetic obese women. Nat Med 24:1070–1080. https://doi.org/10.1038/s41591-018-0061-3

576 19. Mardinoglu A, Wu H, Bjornson E, et al (2018) An Integrated Understanding of the Rapid Metabolic Benefits of a
577 Carbohydrate-Restricted Diet on Hepatic Steatosis in Humans. Cell Metab 27:559-571.e5.
578 https://doi.org/10.1016/j.cmet.2018.01.005

579 20. Loomba R, Seguritan V, Li W, et al (2017) Gut Microbiome-Based Metagenomic Signature for Non-invasive Detection
580 of Advanced Fibrosis in Human Nonalcoholic Fatty Liver Disease. Cell Metab 25:1054-1062.e5.
581 https://doi.org/10.1016/j.cmet.2017.04.001

582 21. Qin N, Yang F, Li A, et al (2014) Alterations of the human gut microbiome in liver cirrhosis. Nature 513:59–64.
583 https://doi.org/10.1038/nature13568

584 22. Lanthier N, Rodriguez J, Nachit M, et al (2021) Microbiota analysis and transient elastography reveal new extra-
585 hepatic components of liver steatosis and fibrosis in obese patients. Sci Rep 11:. https://doi.org/10.1038/s41598-
586 020-79718-9

587 23. Er^ Ome Boursier J, Mueller O, Barret M, et al (2015) The Severity of Nonalcoholic Fatty Liver Disease Is Associated
588 With Gut Dysbiosis and Shift in the Metabolic Function of the Gut Microbiota.
589 https://doi.org/10.1002/hep.28356/suppinfo

590 24. Caussy C, Tripathi A, Humphrey G, et al (2019) A gut microbiome signature for cirrhosis due to nonalcoholic fatty
591 liver disease. Nat Commun 10:. https://doi.org/10.1038/s41467-019-09455-9

592 25. Oh TG, Kim SM, Caussy C, et al (2020) A Universal Gut-Microbiome-Derived Signature Predicts Cirrhosis. Cell Metab
593 32:878-888.e6. https://doi.org/10.1016/j.cmet.2020.06.005

26. Shen F, Zheng RD, Sun XQ, et al (2017) Gut microbiota dysbiosis in patients with non-alcoholic fatty liver disease. Hepatobiliary and Pancreatic Diseases International 16:375–381. https://doi.org/10.1016/S1499-3872(17)60019-5

27. Aron-Wisnewsky J, Vigliotti C, Witjes J, et al (2020) Gut microbiota and human NAFLD: disentangling microbial signatures from metabolic disorders. Nat Rev Gastroenterol Hepatol 17:279–297

28. Segata N, Waldron L, Ballarini A, et al (2012) Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods 9:811–814. https://doi.org/10.1038/nmeth.2066

29. Blanco-Míguez A, Beghini F, Cumbo F, et al (2023) Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. Nat Biotechnol 41:1633–1644. https://doi.org/10.1038/s41587-023-01688-w

30. McPherson S, Hardy T, Dufour JF, et al (2017) Age as a Confounding Factor for the Accurate Non-Invasive Diagnosis of Advanced NAFLD Fibrosis. American Journal of Gastroenterology 112:740–751. https://doi.org/10.1038/ajg.2016.453

31. ponziani20_confounding_NAFLD

32. Chu H, Duan Y, Yang L, Schnabl B (2019) Small metabolites, possible big changes: A microbiota-centered view of non-alcoholic fatty liver disease. Gut 68:359–370. https://doi.org/10.1136/gutjnl-2018-316307

33. Dai X, Hou H, Zhang W, et al (2020) Microbial Metabolites: Critical Regulators in NAFLD. Front Microbiol 11

34. Martin-Grau M, Monleón D (2024) The Role of Microbiota-Related Co-Metabolites in MASLD Progression: A Narrative Review. Curr Issues Mol Biol 46:6377–6389

35. Oliphant K, Allen-Vercoe E (2019) Macronutrient metabolism by the human gut microbiome: Major fermentation by-products and their impact on host health. Microbiome 7:. https://doi.org/10.1186/s40168-019-0704-8

36. Donaldson GP, Lee SM, Mazmanian SK (2015) Gut biogeography of the bacterial microbiota. Nat Rev Microbiol 14:20–32

37. Albillos A, de Gottardi A, Rescigno M (2020) The gut-liver axis in liver disease: Pathophysiological basis for therapy. J Hepatol 72:558–577

38. Iruzubieta P, Medina JM, Fernández-López R, et al (2020) A role for gut microbiome fermentative pathways in fatty liver disease progression. J Clin Med 9

39. Donohoe DR, Garge N, Zhang X, et al (2011) The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. Cell Metab 13:517–526. https://doi.org/10.1016/j.cmet.2011.02.018

40. Wang HB, Wang PY, Wang X, et al (2012) Butyrate enhances intestinal epithelial barrier function via up-regulation of tight junction protein claudin-1 transcription. Dig Dis Sci 57:3126–3135. https://doi.org/10.1007/s10620-012-2259-4

41. Gao Z, Yin J, Zhang J, et al (2009) Butyrate improves insulin sensitivity and increases energy expenditure in mice. Diabetes 58:1509–1517. https://doi.org/10.2337/db08-1637

42. Li Z, Yi CX, Katiraei S, et al (2018) Butyrate reduces appetite and activates brown adipose tissue via the gut-brain neural circuit. Gut 67:1269–1279. https://doi.org/10.1136/gutjnl-2017-314050

43. De Vadder F, Kovatcheva-Datchary P, Goncalves D, et al (2014) Microbiota-generated metabolites promote metabolic benefits via gut-brain neural circuits. Cell 156:84–96. https://doi.org/10.1016/j.cell.2013.12.016

631 44. Sanna S, van Zuydam NR, Mahajan A, et al (2019) Causal relationships among the gut microbiome, short-chain fatty
632 acids and metabolic diseases. Nat Genet 51:600–605

633 45. Endo H, Niioka M, Kobayashi N, et al (2013) Butyrate-Producing Probiotics Reduce Nonalcoholic Fatty Liver Disease
634 Progression in Rats: New Insight into the Probiotics for the Gut-Liver Axis. PLoS One 8:.
635 https://doi.org/10.1371/journal.pone.0063388

636 46. Jin CJ, Sellmann C, Engstler AJ, et al (2015) Supplementation of sodium butyrate protects mice from the development
637 of non-alcoholic steatohepatitis (NASH). British Journal of Nutrition 114:1745–1755.
638 https://doi.org/10.1017/S0007114515003621

639 47. Zhou D, Chen YW, Zhao ZH, et al (2018) Sodium butyrate reduces high-fat diet-induced non-alcoholic steatohepatitis
640 through upregulation of hepatic GLP-1R expression. Exp Mol Med 50:. https://doi.org/10.1038/s12276-018-0183-1

641 48. Zhou D, Pan Q, Xin FZ, et al (2017) Sodium butyrate attenuates high-fat diet-induced steatohepatitis in mice by
642 improving gut microbiota and gastrointestinal barrier. World J Gastroenterol 23:60–75.
643 https://doi.org/10.3748/wjg.v23.i1.60

644 49. Zhou D, Pan Q, Shen F, et al (2017) Total fecal microbiota transplantation alleviates high-fat diet-induced
645 steatohepatitis in mice via beneficial regulation of gut microbiota. Sci Rep 7:. https://doi.org/10.1038/s41598-017-
646 01751-y

647 50. Louis P, Flint HJ (2017) Formation of propionate and butyrate by the human colonic microbiota. Environ Microbiol
648 19:29–41

649 51. Zhu L, Baker SS, Gill C, et al (2013) Characterization of gut microbiomes in nonalcoholic steatohepatitis (NASH)
650 patients: A connection between endogenous alcohol and NASH. Hepatology 57:601–609.
651 https://doi.org/10.1002/hep.26093

652 52. Michail S, Lin M, Frey MR, et al (2015) Altered gut microbial energy and metabolism in children with non-alcoholic
653 fatty liver disease. FEMS Microbiol Ecol 91:. https://doi.org/10.1093/femsec/fiu002

654 53. Meijnikman AS, Davids M, Herrema H, et al (2022) Microbiome-derived ethanol in nonalcoholic fatty liver disease.
655 Nat Med 28:2100–2106. https://doi.org/10.1038/s41591-022-02016-6

656 54. Yuan J, Chen C, Cui J, et al (2019) Fatty Liver Disease Caused by High-Alcohol-Producing Klebsiella pneumoniae. Cell
657 Metab 30:675-688.e7. https://doi.org/10.1016/j.cmet.2019.08.018

658 55. Chen X, Zhang Z, Li H, et al (2020) Endogenous ethanol produced by intestinal bacteria induces mitochondrial
659 dysfunction in non-alcoholic fatty liver disease. Journal of Gastroenterology and Hepatology (Australia) 35:2009–
660 2019. https://doi.org/10.1111/jgh.15027

661 56. Fromenty B, Roden M (2023) Mitochondrial alterations in fatty liver diseases. J Hepatol 78:415–429

662 57. Ruuskanen MO, Åberg F, Männistö V, et al (2021) Links between gut microbiome composition and fatty liver disease
663 in a large population sample. Gut Microbes 13:1–22. https://doi.org/10.1080/19490976.2021.1888673

664 58. Pony P, Rapisarda C, Terradot L, et al (2020) Filamentation of the bacterial bi-functional alcohol/aldehyde
665 dehydrogenase AdhE is essential for substrate channeling and enzymatic regulation. Nat Commun 11:.
666 https://doi.org/10.1038/s41467-020-15214-y

667 59. Hitschler L, Nissen LS, Kuntz M, Basen M (2021) Alcohol dehydrogenases AdhE and AdhB with broad substrate ranges
668 are important enzymes for organic acid reduction in Thermoanaerobacter sp. strain X514. Biotechnol Biofuels 14:.
669 https://doi.org/10.1186/s13068-021-02038-1

670 60. Pick A, Rühmann B, Schmid J, Sieber V (2013) Novel CAD-like enzymes from Escherichia coli K-12 as additional tools
671 in chemical production. Appl Microbiol Biotechnol 97:5815–5824. https://doi.org/10.1007/s00253-012-4474-5

672 61. Merchel Piovesan Pereira B, Adil Salim M, Rai N, Tagkopoulos I (2021) Tolerance to Glutaraldehyde in Escherichia
673 coli Mediated by Overexpression of the Aldehyde Reductase YqhD by YqhC. Front Microbiol 12:.
674 https://doi.org/10.3389/fmicb.2021.680553

675 62. Moreira de Gouveia MI, Daniel J, Garrivier A, et al (2023) Diversity of ethanolamine utilization by human commensal
676 Escherichia coli. Res Microbiol 174:. https://doi.org/10.1016/j.resmic.2022.103989

677 63. Dellomonaco C, Clomburg JM, Miller EN, Gonzalez R (2011) Engineered reversal of the β-oxidation cycle for the
678 synthesis of fuels and chemicals. Nature 476:355–359. https://doi.org/10.1038/nature10333

679 64. Eram MS, Ma K (2013) Decarboxylation of pyruvate to acetaldehyde for ethanol production by hyperthermophiles.
680 Biomolecules 3:578–596

681 65. Kwok KH, Weiner H (2005) Isolation and characterization of an aldehyde dehydrogenase encoded by the aldB gene
682 of Escherichia coli. J Bacteriol 187:1067–1073. https://doi.org/10.1128/JB.187.3.1067-1073.2005

683 66. Dank A, Zeng Z, Boeren S, et al (2021) Bacterial Microcompartment-Dependent 1,2-Propanediol Utilization of
684 Propionibacterium freudenreichii. Front Microbiol 12:. https://doi.org/10.3389/fmicb.2021.679827

685 67. Chen YM, Liu Y, Zhou RF, et al (2016) Associations of gut-flora-dependent metabolite trimethylamine-N-oxide,
686 betaine and choline with non-alcoholic fatty liver disease in adults. Sci Rep 6:. https://doi.org/10.1038/srep19076

687 68. Flores-Guerrero JL, Post A, van Dijk PR, et al (2021) Circulating trimethylamine-N-oxide is associated with all-cause
688 mortality in subjects with nonalcoholic fatty liver disease. Liver International 41:2371–2382.
689 https://doi.org/10.1111/liv.14963

690 69. Wang Z, Klipfell E, Bennett BJ, et al (2011) Gut flora metabolism of phosphatidylcholine promotes cardiovascular
691 disease. Nature 472:57–65. https://doi.org/10.1038/nature09922

692 70. Roncal C, Martínez-Aguilar E, Orbe J, et al (2019) Trimethylamine-N-Oxide (TMAO) Predicts Cardiovascular Mortality
693 in Peripheral Artery Disease. Sci Rep 9:. https://doi.org/10.1038/s41598-019-52082-z

694 71. Martínez-del Campo A, Bodea S, Hamer HA, et al (2015) Characterization and detection of a widely distributed gene
695 cluster that predicts anaerobic choline utilization by human gut bacteria. mBio 6:.
696 https://doi.org/10.1128/mBio.00042-15

697 72. Goris M, Cea-Rama I, Puntervoll P, et al (2023) Increased Thermostability of an Engineered Flavin-Containing
698 Monooxygenase to Remediate Trimethylamine in Fish Protein Hydrolysates. Appl Environ Microbiol 89:e0039023.
699 https://doi.org/10.1128/aem.00390-23

700 73. Leimkühler S (2020) The biosynthesis of the molybdenum cofactors in <scp> *Escherichia coli* </scp>. Environ
701 Microbiol 22:2007–2026. https://doi.org/10.1111/1462-2920.15003

702 74. Jay Ferguson D, Adrian Krzycki J (1997) Reconstitution of Trimethylamine-Dependent Coenzyme M Methylation with
703 the Trimethylamine Corrinoid Protein and the Isozymes of Methyltransferase II from Methanosarcina barkeri

704 75. Ferguson DJ, Krzycki JA, Grahame DA (1996) Specific roles of methylcobamide:coenzyme M methyltransferase
705 isozymes in metabolism of methanol and methylamines in Methanosarcina barkeri. Journal of Biological Chemistry
706 271:5189–5194. https://doi.org/10.1074/jbc.271.9.5189

707 76. Kurth JM, Nobu MK, Tamaki H, et al (2021) Methanogenic archaea use a bacteria-like methyltransferase system to
708 demethoxylate aromatic compounds. ISME Journal 15:3549–3565. https://doi.org/10.1038/s41396-021-01025-6

709 77. Gendron A, Allen KD (2022) Overview of Diverse Methyl/Alkyl-Coenzyme M Reductases and Considerations for Their
710 Potential Heterologous Expression. Front Microbiol 13:. https://doi.org/10.3389/fmicb.2022.867342

711 78. Thauer RK (2019) Methyl (Alkyl)-Coenzyme M Reductases: Nickel F-430-Containing Enzymes Involved in Anaerobic
712 Methane Formation and in Anaerobic Oxidation of Methane or of Short Chain Alkanes. Biochemistry 58:5198–5220.
713 https://doi.org/10.1021/acs.biochem.9b00164

714 79. Wagner T, Koch J, Ermler U, Shima S (2017) Methanogenic heterodisulfide reductase (HdrABC-MvhAGD) uses two
715 noncubane [4Fe-4S] clusters for reduction. Science (1979) 357:699–703. https://doi.org/10.1126/science.aan0425

716 80. Aziz I, Kayastha K, Kaltwasser S, et al (2024) Structural and mechanistic basis of the central energy-converting
717 methyltransferase complex of methanogenesis. Proc Natl Acad Sci U S A 121:.
718 https://doi.org/10.1073/pnas.2315568121

719 81. Almeida A, Nayfach S, Boland M, et al (2021) A unified catalog of 204,938 reference genomes from the human gut
720 microbiome. Nat Biotechnol 39:105–114. https://doi.org/10.1038/s41587-020-0603-3

721 82. Rosconi F, Rudmann E, Li J, et al (2022) A bacterial pan-genome makes gene essentiality strain-dependent and
722 evolvable. Nat Microbiol 7:1580–1592. https://doi.org/10.1038/s41564-022-01208-7

723 83. Yu MK, Fogarty EC, Eren AM (2024) Diverse plasmid systems and their ecology across human gut metagenomes
724 revealed by PlasX and MobMess. Nat Microbiol 9:830–847. https://doi.org/10.1038/s41564-024-01610-3

725 84. Medina JM, Fernández-López R, Crespo J, Cruz F de la (2021) Propionate fermentative genes of the gut microbiome
726 decrease in inflammatory bowel disease. J Clin Med 10:. https://doi.org/10.3390/jcm10102176

727 85. Louis P, McCrae SI, Charrier C, Flint HJ (2007) Organization of butyrate synthetic genes in human colonic bacteria:
728 Phylogenetic conservation and horizontal gene transfer. FEMS Microbiol Lett 269:240–247.
729 https://doi.org/10.1111/j.1574-6968.2006.00629.x

730 86. Cui Y, Ma X, Lee SH, et al (2023) Production of butyl butyrate from lignocellulosic biomass through Escherichia coli-
731 Clostridium beijerinckii G117 co-culture. Process Biochemistry 128:58–67.
732 https://doi.org/10.1016/j.procbio.2023.02.005

733 87. Gonzalez-Alba JM, Baquero F, Cantón R, Galán JC (2019) Stratified reconstruction of ancestral Escherichia coli
734 diversification. BMC Genomics 20:. https://doi.org/10.1186/s12864-019-6346-1

735 88. Palomino A, Gewurz D, DeVine L, et al (2023) Metabolic genes on conjugative plasmids are highly prevalent in
736 Escherichia coli and can protect against antibiotic treatment. ISME Journal 17:151–162.
737 https://doi.org/10.1038/s41396-022-01329-1

738 89. Abriouel H, Pérez Montoro B, de la Fuente Ordoñez JJ, et al (2019) New insights into the role of plasmids from
739 probiotic Lactobacillus pentosus MP-10 in Aloreña table olive brine fermentation. Sci Rep 9:.
740 https://doi.org/10.1038/s41598-019-47384-1

741 90. Pudlo NA, Urs K, Crawford R, et al (2022) Phenotypic and Genomic Diversification in Complex Carbohydrate-
742 Degrading Human Gut Bacteria

743 91. Bushnell B (2012) BBTools: https://sourceforge.net/projects/bbtools/

744 92. Andrews S (2010) FastQC: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

745 93. Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: Summarize analysis results for multiple tools and samples
746 in a single report. Bioinformatics 32:3047–3048. https://doi.org/10.1093/bioinformatics/btw354

747 94. Caspi R, Billington R, Fulcher CA, et al (2018) The MetaCyc database of metabolic pathways and enzymes. Nucleic
748 Acids Res 46:D633–D639. https://doi.org/10.1093/nar/gkx935

749 95. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance
750 and usability. Mol Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010

751 96. Minh BQ, Schmidt HA, Chernomor O, et al (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
752 Inference in the Genomic Era. Mol Biol Evol 37:1530–1534. https://doi.org/10.1093/molbev/msaa015

753 97. Eddy S (2023) HMMER: biosequence analysis using profile hidden Markov models

754 98. Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes

755 99. Buchfink B, Reuter K, Drost HG (2021) Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat
756 Methods 18:366–368. https://doi.org/10.1038/s41592-021-01101-x

757 100. Treiber ML, Taft DH, Korf I, et al (2020) Pre-and post-sequencing recommendations for functional annotation of
758 human fecal metagenomes. BMC Bioinformatics 21:. https://doi.org/10.1186/s12859-020-3416-y

759 101. O'Leary NA, Wright MW, Brister JR, et al (2016) Reference sequence (RefSeq) database at NCBI: Current status,
760 taxonomic expansion, and functional annotation. Nucleic Acids Res 44:D733–D745.
761 https://doi.org/10.1093/nar/gkv1189

762 102. Hyatt D, Chen GL, LoCascio PF, et al (2010) Prodigal: Prokaryotic gene recognition and translation initiation site
763 identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119

764 103. Wickham H, Averick M, Bryan J, et al (2019) Welcome to the Tidyverse. J Open Source Softw 4:1686.
765 https://doi.org/10.21105/joss.01686

766 104. Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis, 2nd ed. Springer Cham

767 105. Kassambara A (2023) ggpubr: "ggplot2" Based Publication Ready Plots

768 106. Kolde R (2019) pheatmap: Pretty Heatmaps

769 107. Kassambara A (2023) rstatix: Pipe-Friendly Framework for Basic Statistical Tests

770