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 (MGEs) 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 dysfunction-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 metabolic dysfunction-associated steatohepatitis (MASH), which can further progress to fibrosis, cirrhosis and hepatocellular carcinoma[3]. MASLD also significantly increases

42 the risk of cardiovascular disease, extrahepatic malignancies (e.g., colorectal or breast cancer), liver-
43 related 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 the
54 bile acid composition and circulation affects liver metabolism and function. Because they play a role in
55 the regulation of glucose and lipid metabolism, their dysregulation contributes to gut inflammation and
56 MASLD[10,11].

57 Several studies have tried to establish a causal link between bacterial metabolism and
58 MASLD[12,13]. However, metabolic levels can also be influenced by host metabolism, diet or
59 environmental factors. Similarly, metataxonomic analysis have explored the relationship between GM
60 composition and MASLD[14,15]. Although 16S rRNA gene sequencing provides taxonomic resolution at the
61 genus or species level, it fails to capture strain-level metabolic variability. Bacterial genomic plasticity
62 allows key metabolic genes, such as those potentially relevant to MASLD pathogenesis, to be encoded
63 in their accessory genome, including MGEs, further complicating the association between taxonomy

64 and metabolic function[16]. In this study, we used the genes encoding metabolic enzymes involved in the
65 final steps of microbial pathways leading to the production of butyrate, short-chain alcohols (SCAs)
66 ethanol and propanol, methane trimethylamine (TMA) and trimethylamine N-oxide (TMAO) as proxies
67 to determine their abundance in MASLD. To achieve this, we isolated these gene families and quantified
68 them in over 550 whole metagenomes obtained from fecal samples of MASLD and healthy patients,
69 stratified across three cohorts. We also explored taxonomic signatures to gain a comprehensive view
70 of bacterial diversity and abundance. Finally, we assessed the presence of these genes within the
71 human GM genomes and plasmids. This approach provided insights into the role of microbial pathways
72 in MASLD progression and highlighted key genetic signatures in both healthy and MASLD patients.

73 **2. Results**

74 ***Agathobacter rectalis* is depleted in MASLD**

75 Certain bacterial clades have been positively correlated with MASLD progression, including
76 *Enterobacteriaceae*[15,17] such as *Escherichia / Shigella spp*.[17,18], *Bacteroidaceae* like *Bacteroides spp.*[18,19],
77 *Veillonellaceae*[15] or *Streptococcus spp.*[20]. Conversely, other clades have been reported to exhibit an
78 inverse correlation with MASLD progression, including *Ruminococcaceae*[15] such as *Faecalibacterium*
79 *prausnitzii*[18,21], *Lachnospiraceae* like *Eubacterium rectale*[18,21] or *Dorea longicatena*[21], *Clostridiaceae* like
80 *Clostridium spp.*[17], or even *Pseudomonas spp.*[20]. However, these studies often yield conflicting results.
81 For example, a reported a positive correlation between *Ruminoccocus spp.* abundance and MASLD
82 progression[19] contradicts one of the main findings in two posterior studies[18,21]. Similarly, *Prevotella spp.*
83 was found to decrease with MASLD development[22] and fibrosis stage[19], yet another study reported an
84 increase in this genus in MASLD patients[20]. Additional examples of discordant taxonomic signatures in
85 MASLD, such as *Blautia spp.* or *Roseburia spp.,* are reviewed elsewhere[23].

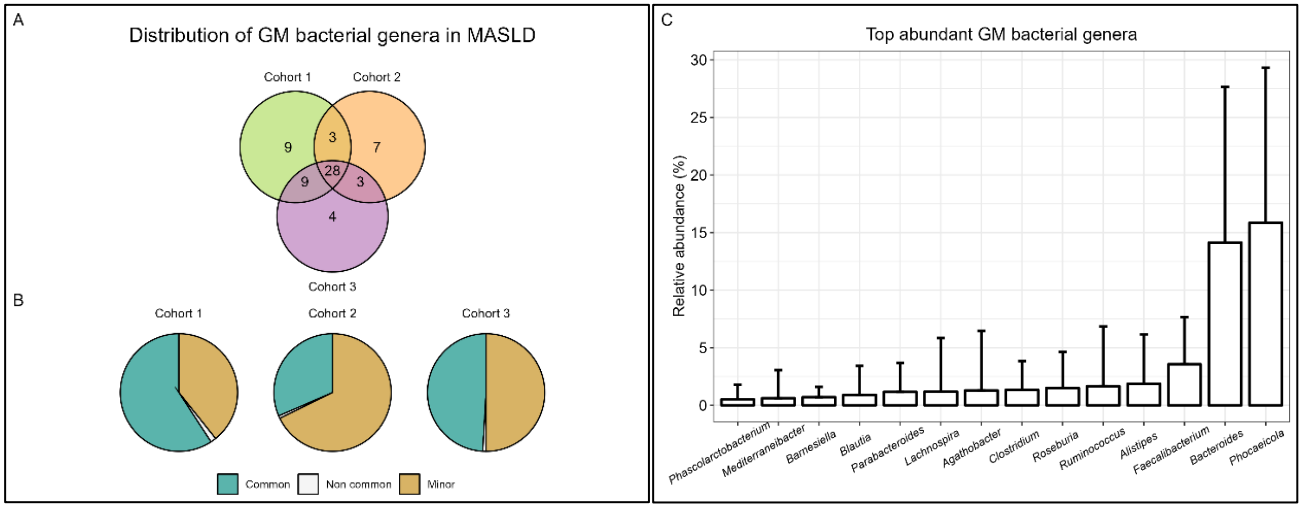
86 To investigate whether any bacterial genera is systematically altered in MASLD, we profiled the
87 taxonomic composition of the GM communities across all samples from the three cohorts using
88 MetaPhlAn[24], as detailed in Materials and Methods. This analysis identified 28 bacterial genera shared
89 across all three cohorts (Fig. 1A) that accounted for 59%, 31.5% and 49% of the GM fraction in each
90 cohort, respectively (Fig. 1B). Thus, nearly half of the bacterial genera identified in at least one cohort
91 were shared across all cohorts (28 out of 63), comprising between one and two-thirds of their total GM
92 abundance.

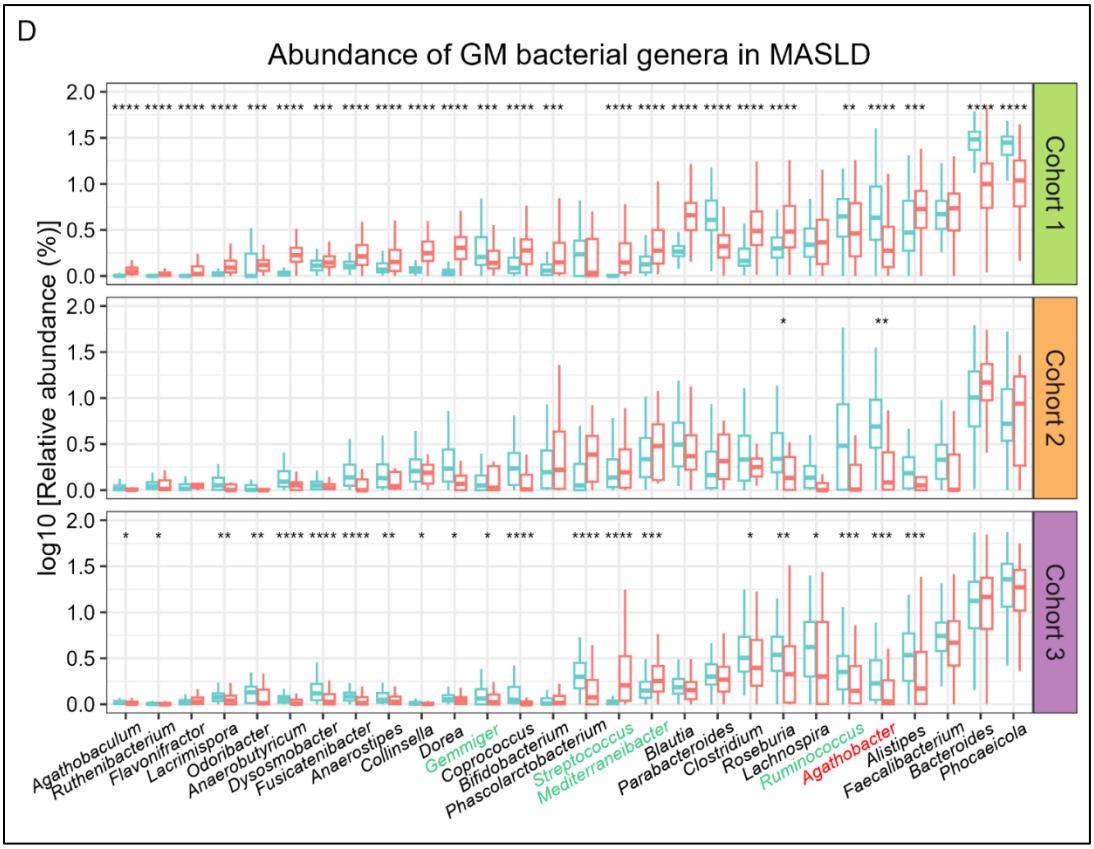
93 Among these shared genera, *Gemmiger*, *Streptococcus*, *Mediterraneibacter*, *Ruminococcus* and
94 *Agathobacter* exhibited a consistent differential abundance pattern between groups across every
95 cohort statistically significant in at least two cohorts (Fig. 1D). This exclusion criterion was applied to
96 enhance the reliability of the results, ensuring that differentially abundant genera were not solely
97 driven by heterogeneous demographic variables (e.g. geographical location, age, sex, or ethnicity) or
98 environmental factors (e.g. diet, physical activity, comorbidities, or drug exposure), rather than being
99 directly related to the MASLD phenotype[25,26]. However, this threshold was not applied to Cohort 2
100 because all patients in this group have MASLD, and applying it could have been overly restrictive,
101 masking potential taxonomic differences associated with the disease. From these five genera, only
102 *Agathobacter* spp. showed statistically significant differences across the three cohorts (Fig. 1D). Note
103 that *Agathobacter*, as well as *Ruminococcocus* and *Mediterraneibacter*, belong to the fraction of the
104 highest abundant GM bacterial genera (Fig. 1C), highlighting their potential contribution to the
105 phenotype in specific patient cohorts.

106 Analogously, the highest abundant bacterial species within the GM were interrogated for
107 differences in abundances between groups across the three cohorts (Fig. 1E). Only *Agathobacter*

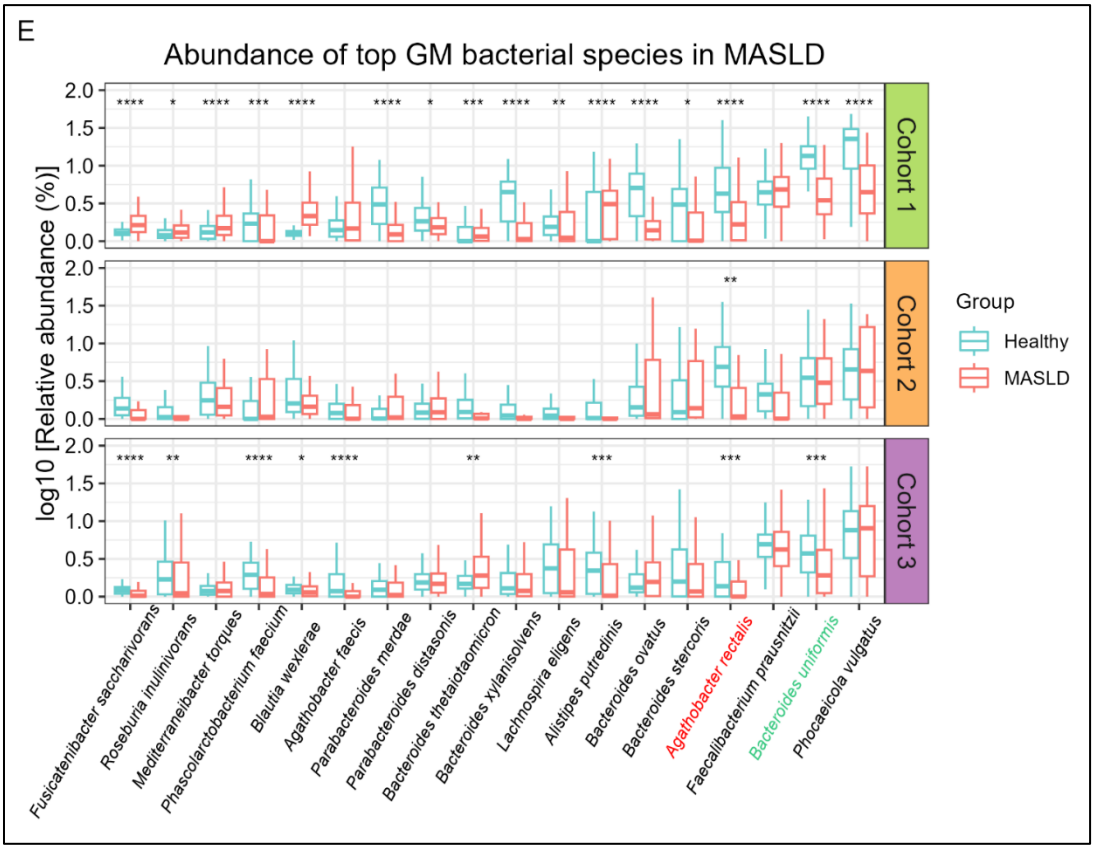
rectalis and *Bacteroides uniformis*, two of the most abundant GM bacterial species (Supplementary Fig. 1), exhibited a consistent pattern of differential abundance across all cohorts, with statistical significance in at least two. However, only *A. rectalis* showed a significant difference between groups in all cohorts, identifying it as a robust taxonomic marker depleted in MASLD.

Figure 1 (cont. next page). Taxonomic profiling of GM 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. The blue section of the pie charts reflects the average relative abundance of bacterial genera common to the three cohorts (n=28), 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)** Top abundant GM genera (average absolute abundance >0.5%) **(D-E)** Relative abundance of bacterial genera **(D)** and species **(E)** 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. Clades depicted in green across the x-axis have a consistent differential abundance pattern between groups over the three cohorts. Only *Agathobacter rectalis* 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.




D
Abundance of GM bacterial genera in MASLD
log10 [Relative abundance (%)]
Cohort 1
Cohort 2
Cohort 3
Agathobaculum
Ruthenibacterium
Flavonifractor
Lacrimispora
Odoribacter
Anaerobutyricum
Dysosmobacter
Fusicatenibacter
Anaerostipes
Collinsella
Dorea
Gemmiger
Coprococcus
Bifidobacterium
Phascolarctobacterium
Streptococcus
Mediterraneibacter
Blautia
Parabacteroides
Clostridium
Roseburia
Lachnospira
Ruminococcus
Agathobacter
Alistipes
Faecalibacterium
Bacteroides
Phocaeicola

128


E
Abundance of top GM bacterial species in MASLD
log10 [Relative abundance (%)]
Cohort 1
Cohort 2
Cohort 3
Group
Healthy
MASLD
Fusicatenibacter saccharivorans
Roseburia inulinivorans
Mediterraneibacter torques
Phascolarctobacterium faecium
Blautia wexlerae
Agathobacter faecis
Parabacteroides merdae
Parabacteroides distasonis
Bacteroides thetaiotaomicron
Bacteroides xylanisolvens
Lachnospira eligens
Alistipes putredinis
Bacteroides ovatus
Bacteroides stercoris
Agathobacter rectalis
Faecalibacterium prausnitzii
Bacteroides uniformis
Phocaeicola vulgatus

129

130 **Butyrate-producing genes are depleted in MASLD**

131 The identification of *A. rectalis* as a consistent taxonomic signature across multiple MASLD
132 cohorts underscores its potential protective role in MASLD development. However, taxonomy alone
133 may not fully capture GM functional contributions. Certain microbial metabolites regulate GLA integrity
134 and have been suggested to influence MASLD progression[27–29]. These compounds are catabolized from
135 dietary carbohydrates and proteins that escape upper gastrointestinal digestion by the GM through
136 fermentative pathways that are predominant in the anaerobic environment of the colon[30,31]. Once
137 absorbed into the bloodstream, they are transported to the liver via the portal vein, where they can
138 exert both beneficial and detrimental effects[32,33].

139 The most well-characterized metabolic family are short chain fatty acids, which are produced
140 by the GM through the colonic fermentation of complex carbohydrates present in dietary fiber that
141 escape digestion in the small intestine. These metabolic pathways have been relatively well understood
142 for nearly thirty years[34] and their impact on GM homeostasis has been reviewed in several articles[35–37].
143 For this reason, we interrogated the metagenomic fecal samples from the patient cohorts for metabolic
144 signatures that could explain the MASLD phenotype. First, we focused on the genes encoding enzymes
145 involved in butyrate formation as a proxy for the bacteria that carry them, independently of their taxa.
146 To validate our hypothesis, we analyzed the genic abundance of the final enzymes in butyrate
147 production.

148 We selected butyrate as a target because it is crucial for maintaining GLA homeostasis by
149 supporting the colonic epithelial barrier[38,39]. It also promotes lipid oxidation[40,41], induces hepatic
150 gluconeogenesis[42], and prevents insulin resistance[43], all of which are key factors in MASLD. In mice,
151 both butyrate-producing probiotics[44] and sodium butyrate administration[45–47] have been shown to

152 prevent steatosis and slow MASLD progression. Additionally, fecal microbiota transplantation (FMT)
153 from healthy to MASH mice increased cecal butyrate levels and reduced steatosis[48].

154 Bacterial butyrate can be formed through two different terminal reactions (Fig. 2A). The first is
155 a dephosphorylation of butyryl-P that yields one ATP through an energetically favorable reaction that
156 is catalyzed by a butyrate kinase encoded by *buk.* In the precursor reaction, butyryl-P is formed through
157 a phosphorylation of butyryl-CoA catalyzed by a phosphate butyryl transferase encoded by *ptb*.
158 Butyrate also accumulates after the transference of a CoA cofactor from butyryl-CoA to acetate through
159 butyryl-CoA:acetate CoA transferase, encoded by *but*. This CoA-transferase route also conserves the
160 energy of the CoA bond in the newly formed CoA-moiety of the co-substrate. Butyryl-CoA is produced
161 from crotonyl-CoA through two different enzymes: butyryl--CoA trans-2-oxidorreductase, encoded by
162 *fabV*, and butyryl--CoA dehydrogenase, encoded by *bcd.* This pathway is further reviewed in the human
163 GM elsewhere[49].

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

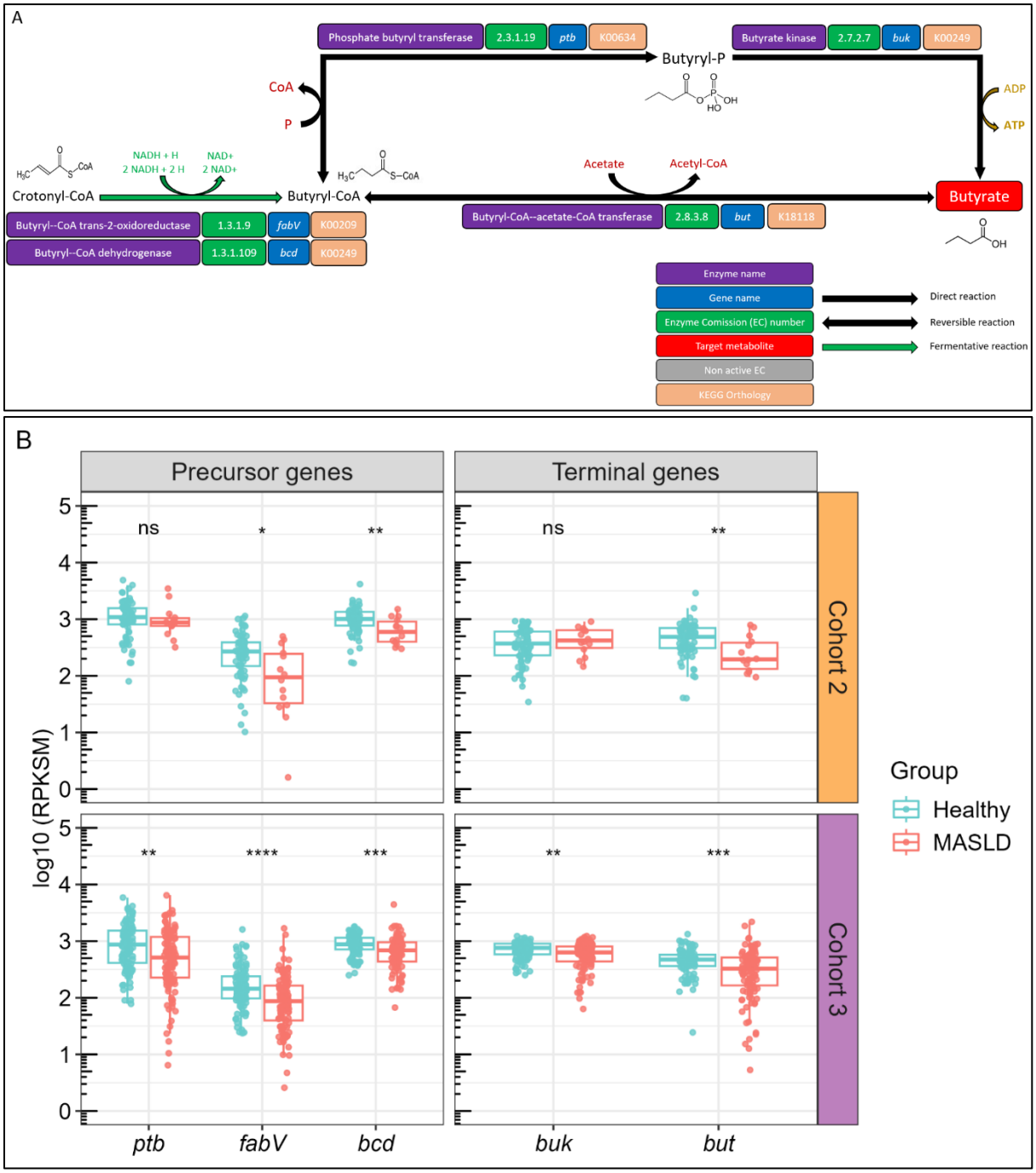


Figure 2. Abundance of butyrate-producing genes in patient cohorts. (A) Characterization of the terminal reactions involved in the formation of bacterial butyrate from crotonyl-CoA. Enzymes, EC numbers, coding genes and KOs, as well as the reaction type, are indicated for each metabolic step according 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 fecal samples, and differences in abundances were evaluated using a pairwise Mann-Whitney test with a Benjamini-Hochberg adjustment for multiple testing. Upper and lower facets represent abundances in Cohorts 2 and 3, respectively. Left and right facets represent abundances of precursor and terminal coding genes, respectively, as described in (A). RPKSM: reads per kilobase per genic size per million reads. (*) $p<0.05$, (**) $p<0.01$, (***) $p<0.001$, (****) $p<0.0001$, ns: non-significant.

185 **SCA-producing genes are increased in MASLD**

186 SCAs are a metabolic family also suggested to regulate GLA integrity, with several studies linking
187 GM endogenous alcohol production to MASLD. The first, published over a decade ago, reported
188 elevated peripheral blood ethanol and fecal *Escherichia* spp. 16S-rRNA levels in MASH patients[50].
189 Similar findings were reported in a pediatric cohort, where children with MASLD exhibited higher levels
190 of fecal ethanol, *Prevotella* spp., and *Gammaproteobacteria* 16S rRNA[51]. A more recent study showed
191 that postprandial peripheral blood ethanol and fecal *Lactobacillaceae* 16S-rRNA levels increased in
192 MASH patients after selective inhibition of host ADH, an effect that was reversed with broad-spectrum
193 antibiotic treatment[52]. In 2019, Yuan and colleagues identified a high-alcohol-producing strain of
194 *Klebsiella pneumoniae* capable of causing MASLD through ethanol production[53]. Specifically, they found
195 that FMT from a MASH patient containing this strain induced MASLD in healthy mice, whereas removing
196 the strain before FMT prevented the disease. Furthermore, ethanol produced by this *K. pneumoniae*
197 strain was shown to disrupt fat homeostasis in hepatic cells and cause mitochondrial dysfunction,
198 oxidative stress, and the accumulation of reactive oxygen species[54], which are critical to MASH
199 development as they drive lipid peroxidation and inflammation[55]. Additionally, a recent computational
200 study identified a positive correlation between the abundance of two fermentative pathways in certain
201 *Clostridia* families and fatty-liver disease in patients. These pathways include the production of ethanol
202 from pyruvate and the heterolactic fermentative pathway[56].

203 With this ample evidence, we hypothesized that genes encoding the enzymes involved in the
204 formation of endogenous SCAs could be another metabolic mechanism involved in MASLD pathology.
205 We focused on ethanol and propanol, which are produced by the reduction of their respective
206 aldehydes through ethanol and propanol dehydrogenases. These enzymes help to reduce the levels of

207 toxic aldehydes in the GM[57,58] and replenish NAD^+ levels to maintain fermentative growth. Bacterial
208 ethanol is formed from acetaldehyde through a fermentative reaction that can be catalyzed by an
209 NADH-dependent, bifunctional acetaldehyde-ethanol dehydrogenase encoded by *adhE*, an NADPH-
210 dependent alcohol dehydrogenase encoded by *adhB*[59,60] and a NADPH-dependent aldehyde reductase
211 encoded by two different genes: *ahr* and *yahK*, with the corresponding liberation of phosphorylated
212 reducing power in form of $NADP^+$ [61,62]. Precursor acetaldehyde can be formed from 1) ethanolamine
213 degradation, through a reaction catalyzed by ethanolamine ammonia-lyase which is encoded by *eutB*
214 and *eutC* genes and liberates ammonia as a by-product[63], 2) acetyl-CoA, through a fermentative
215 reaction catalyzed by bifunctional AdhE and by a CoA-acetylating acetaldehyde dehydrogenase
216 encoded by *mhpF*[64,65] and 3) acetate, through a dehydration catalyzed by aldehyde dehydrogenase B,
217 encoded by *aldB*[66] (Fig. 3A, top). Bacterial 1-propanol is formed from propionaldehyde through a
218 fermentative reaction catalyzed by a propanol dehydrogenase encoded by two different genes: *adhP*
219 (a propanol-preferring alcohol dehydrogenase) and *pduQ.* Propionaldehyde can be generated through
220 two distinct reactions: a CoA transference from propionyl-CoA catalyzed by a propionaldehyde
221 dehydrogenase encoded by *pduP*; and a dehydration of 1,2 propanediol catalyzed by a propanediol
222 dehydratase composed by three subunits encoded by *pduC*, *pduD* and *pduE.* These genes are co-located
223 within the same operon and encode the corresponding enzymes responsible for the metabolism of 1,2-
224 propanediol within the propanediol-utilizing (Pdu) microcompartment, a specialized bacterial structure
225 found in certain enteric species[67,68] (Fig. 3A, bottom).

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

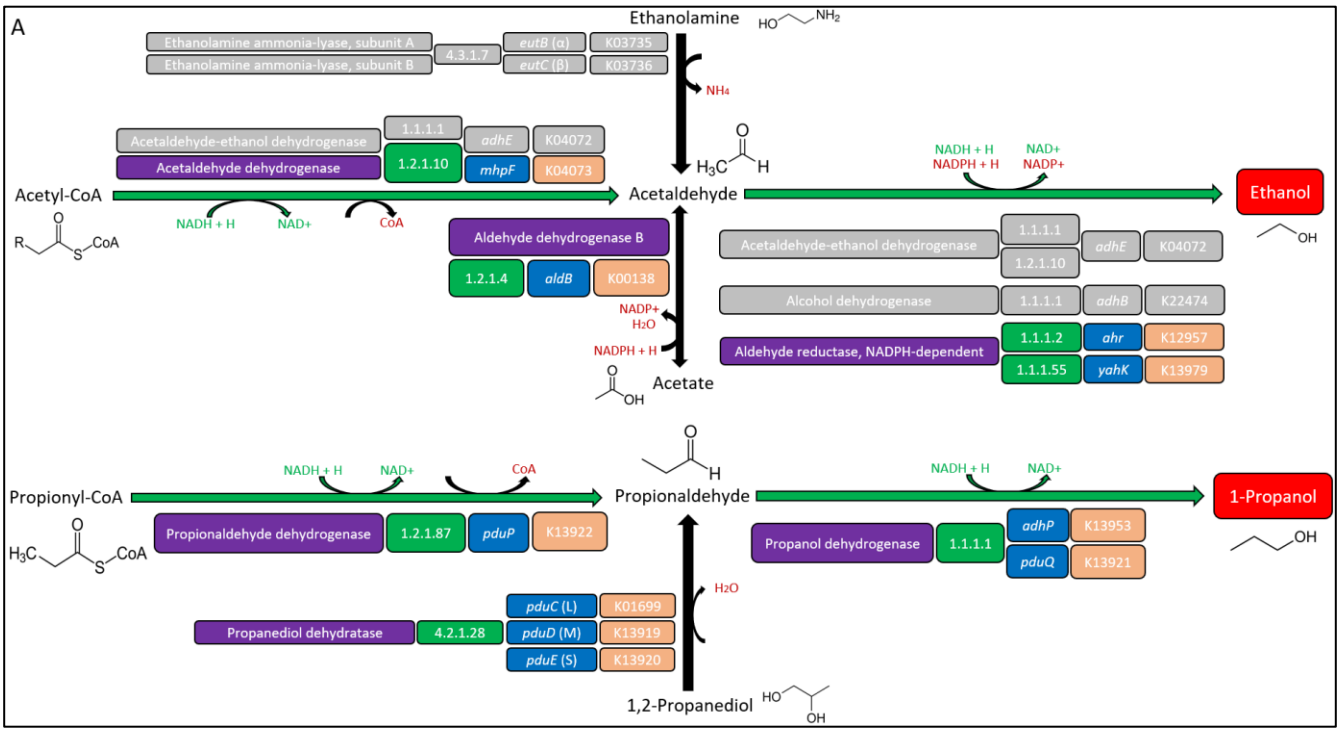
229 patients from Cohort 1 and Cohort 3 compared to healthy groups. Their abundance also increased with

230 MASLD progression (Fig. 3B). *mhpF* and *ahr* were depleted compared to *aldB* and *yahK* respectively in

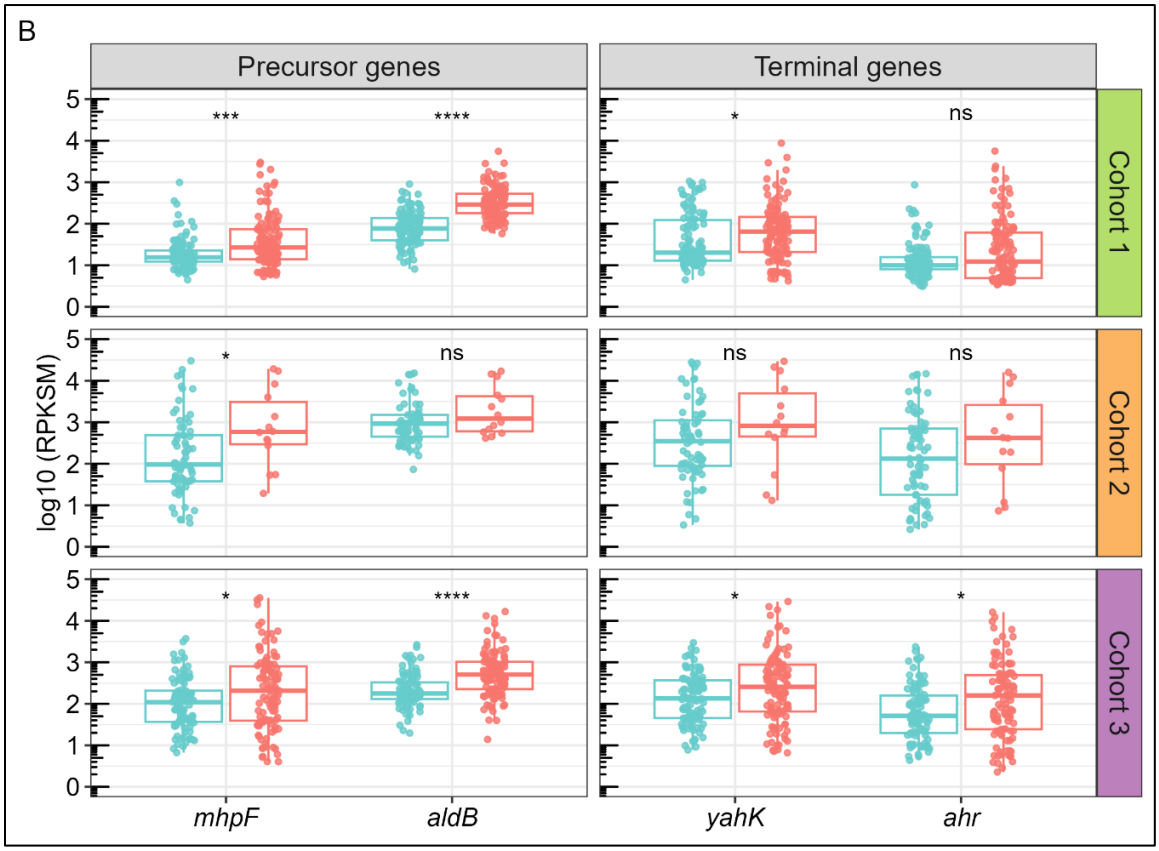
231 all three cohorts. All genes involved in the formation of 1-propanol and propionaldehyde were also

232 significantly enriched in MASLD patients by more than half order of magnitude, increasing with MASLD

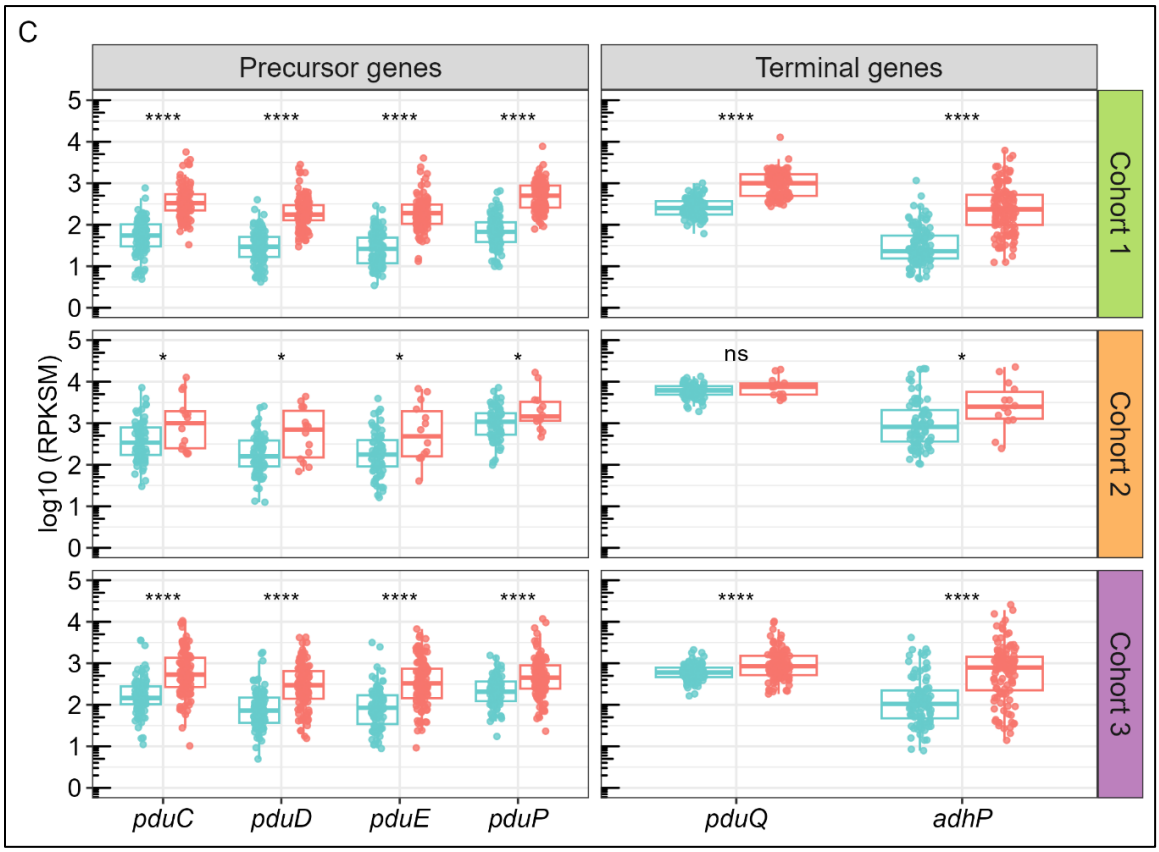
233 progression (Fig. 3C). These findings reveal a systematic increase of SCA-producing genes in MASLD.



234


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

235


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

236

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

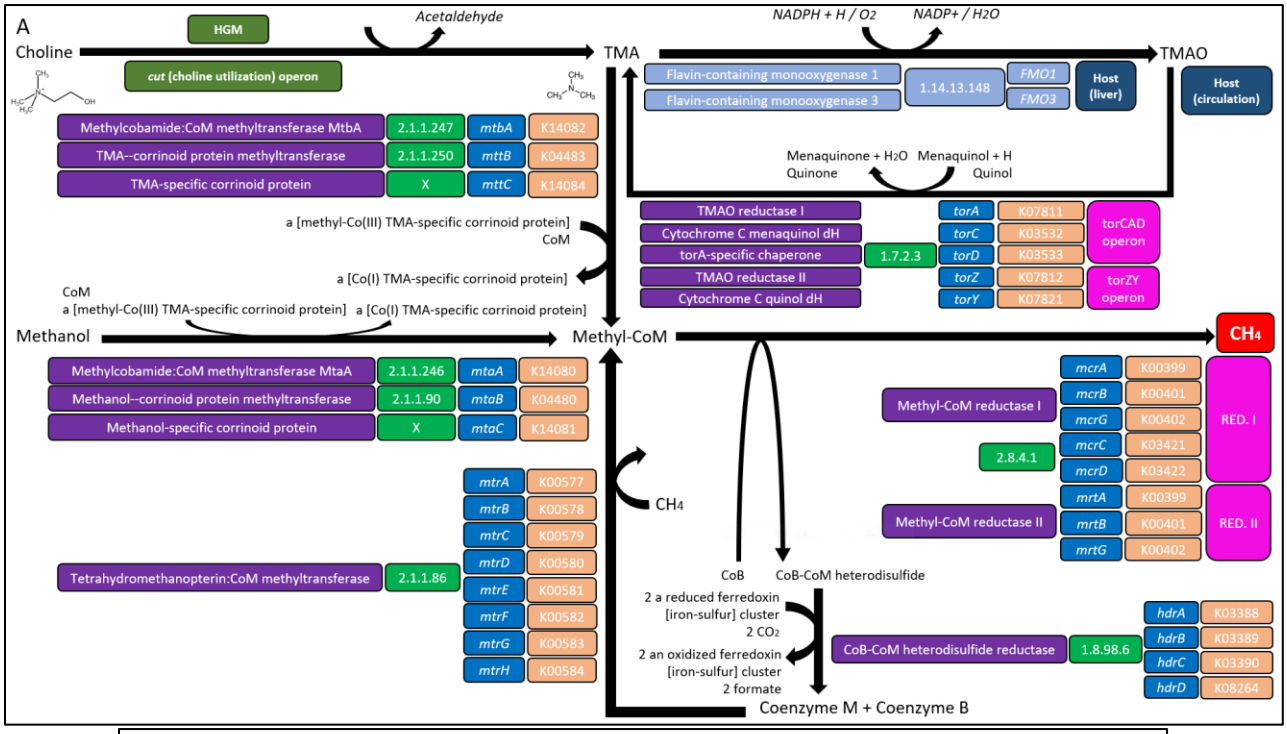
244 **Genes involved in methane production are decreased in MASLD, whereas *tor* operons driving TMA**
245 **accumulation are elevated**

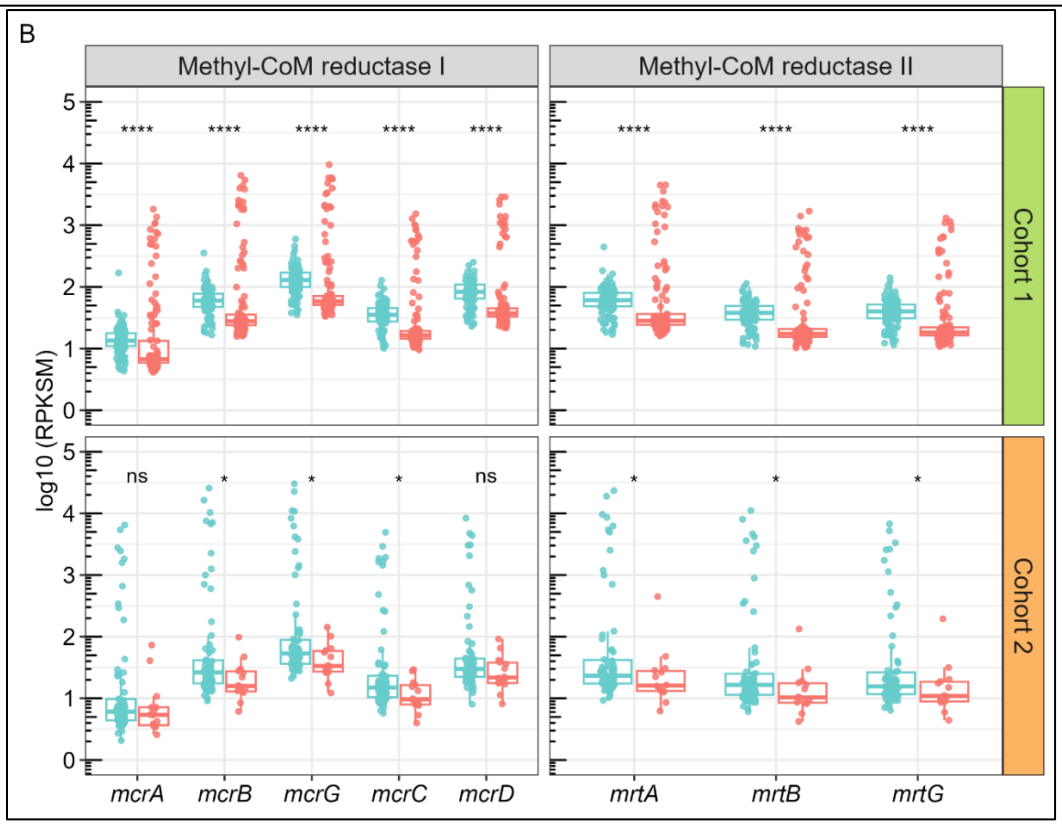
246 Circulating TMAO levels have also been linked to MASLD severity[69] and all-cause mortality in
247 MASLD patients[70]. Higher plasmatic TMAO levels also increase cardiovascular disease risk -a condition
248 associated with MASLD- in both mice[71] and clinical patients[72]. Therefore, we investigated whether
249 bacterial TMAO could be another metabolic mechanism involved in MASLD.

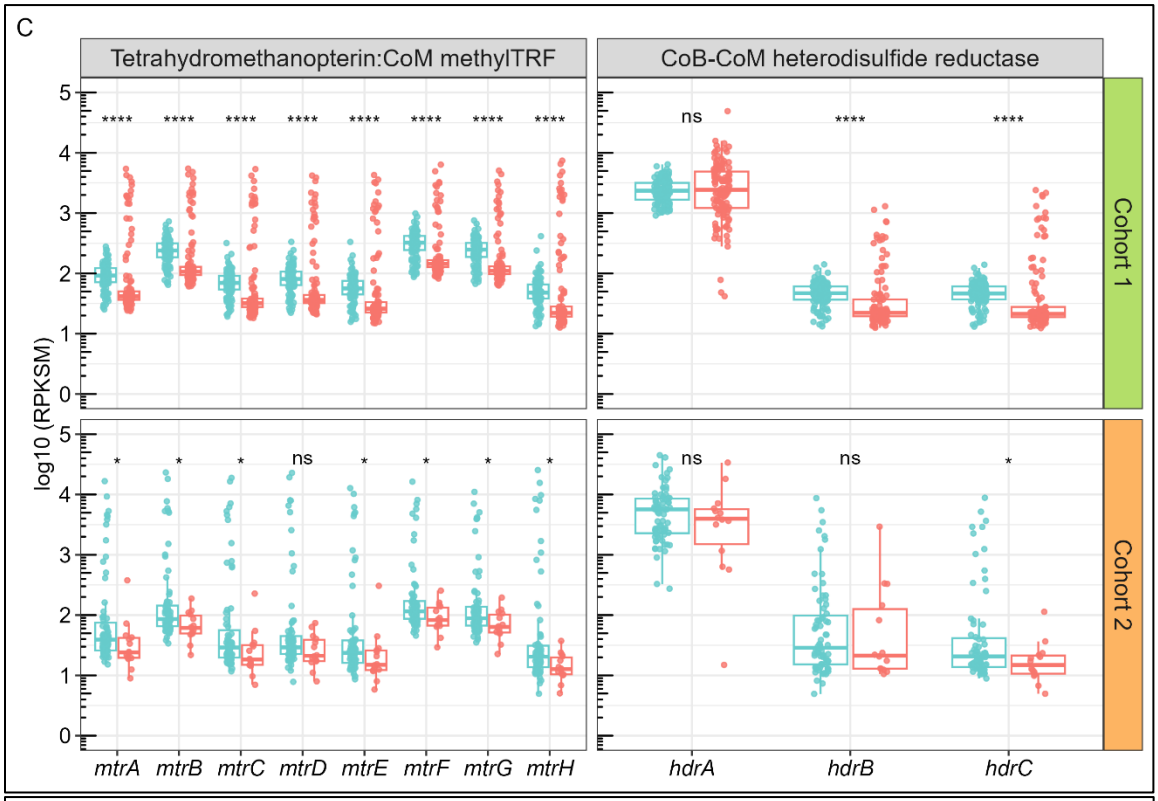
250 Methane production in the GM occurs through a two-step process: a simultaneous TMA
251 demethylation and CoM methylation, followed by the reduction of methyl-CoM (Fig. 4A). TMA
252 originates from choline degradation via the *cut* operon[73] and is oxidized in the liver to TMAO[74] or
253 reduced back to TMA under anaerobic conditions via the *torCAD* and *torZY* operons. Specifically, *torA*
254 and *torZ* encode two TMAO reductases, whose molecular mechanisms are reviewed elsewhere[75]. CoM
255 methylation occurs through two well-characterized enzymatic systems: a TMA-specific (composed of
256 *mttC*, *mtbA* and *mttB*) and a methanol-specific pathway (composed of *mtaC, mtaA* and *mtaB*)[76]. The
257 final reduction of methyl-CoM to methane is catalyzed by methyl-CoM reductase (MCR) I, which is
258 encoded by the *mcrBDCGA* operon[77]. MCR I has an alternative isozyme termed MCR II, encoded by
259 *mrtBGA* operon[78]. The resulting heterodisulfide (CoB-S-S-CoM) is then reduced to CoM and CoB by a
260 heterodisulfide reductase encoded by operon *hdrABC*, which regenerate both cofactors[79]. Additionally,
261 methyl-CoM is replenished through tetrahydromethanopterin CoM-methyltransferase (encoded by
262 operon *mtrABCDEFGH*), which methylates CoM[80].

263 We measured the abundance of these genes in the samples of the three MASLD cohorts as
264 described above. The MCR I and II coding genes, involved in reducing methyl-CoM to methane,
265 significantly decreased in the GM of MASLD patients. Their abundance also declined with MASLD
266 progression (Fig. 4B). Concordantly, *hdr* and *mtr* genes, responsible for replenishing CoM and
267 transferring it to a methyl group to form precursor methyl-CoM, were significantly decreased in MASLD
268 and showed a further decrease with MASLD progression (Fig. 4C). Interestingly perhaps, *hdrA* was two
269 orders of magnitude more abundant than the rest of the genes from the operon. *mtbA, mttB, mtaA*
270 and *mtaB* were significantly decreased in MASLD. However, they did not exhibit differences between
271 the initial and advanced stages of MASLD (Supplementary Fig. 2). Notably, the higher abundance of
272 methane-producing genes observed in the healthy group was not replicated in Cohort 3. In contrast,
273 bacterial genes involved in TMA regeneration showed the opposite trend. The *torCAD* and *torZY*
274 operons were significantly increased in the GM of MASLD patients, and their abundance increased with
275 MASLD progression (Fig. 4D). These findings suggest that methane production from TMA and methanol
276 that naturally occurs in a healthy state is disrupted in MASLD, revealing another potential microbial
277 pathway involved in MASLD pathogenesis.

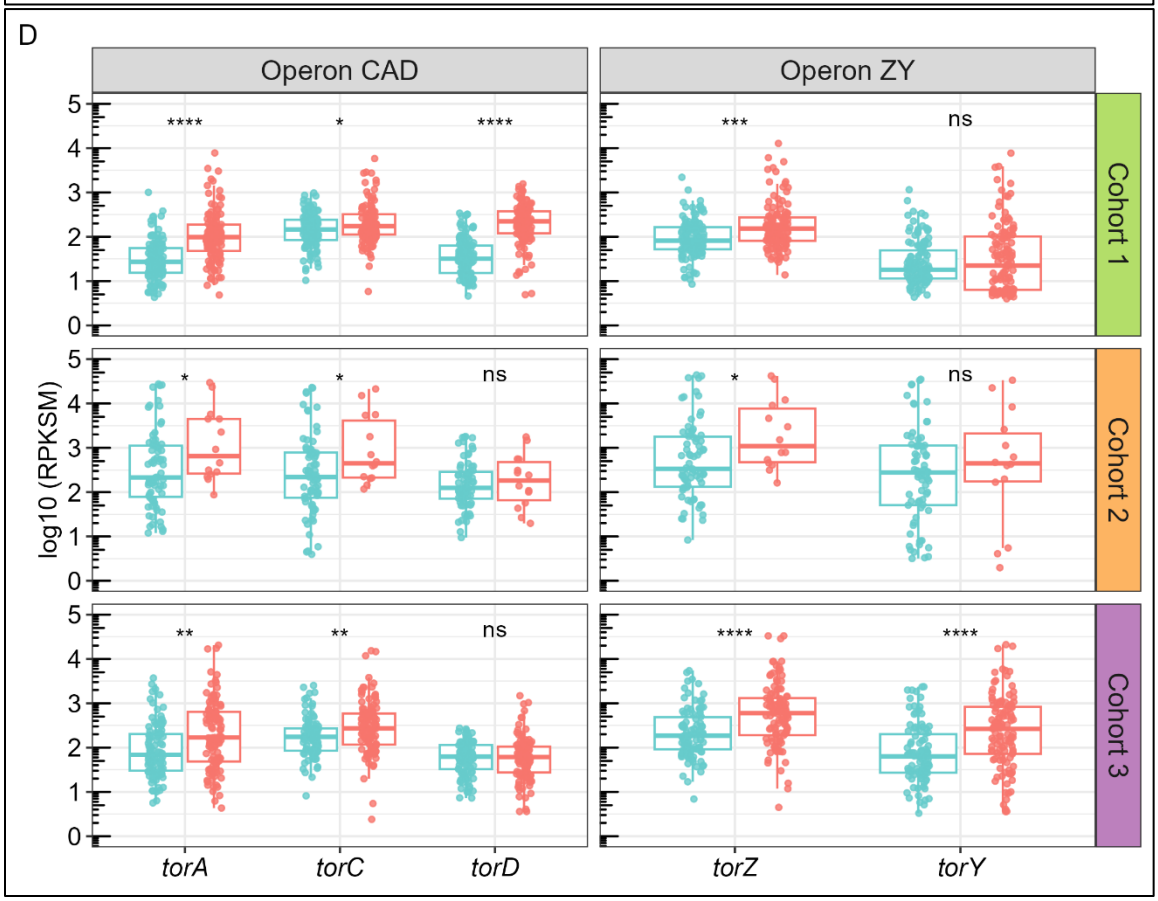
278 Finally, we measured the abundance of five universal single-copy genes (USCGs), as described
279 in Material and Methods, to confirm that observed differences in genic abundance between sample
280 groups were truly due to MASLD phenotype and not influenced by other biological factors nor spurious
281 technical artifacts derived from the quantification method. We found no differences in gene abundance
282 between groups for any USCGs in any of the cohorts (Supplementary Fig. 1).


A
Choline
HGM
Acetaldehyde
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
a [methyl-Co(III) TMA-specific corrinoid protein]
CoM
a [Co(I) TMA-specific corrinoid protein]
TMAO reductase I
Cytochrome C menaquinol dH
torA-specific chaperone
TMAO reductase II
Cytochrome C quinol dH
1.7.2.3
torA K07811
torC K03532
torD K03533
torZ K07812
torY K07821
torCAD operon
torZY operon
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
2.8.4.1
mcrA K00399
mcrB K00401
mcrG K00402
mcrC K03421
mcrD K03422
RED. I
Methyl-CoM reductase II
mrtA K00399
mrtB K00401
mrtG K00402
RED. II
CH4
mtrA K00577
mtrB K00578
mtrC K00579
mtrD K00580
mtrE K00581
mtrF K00582
mtrG K00583
mtrH K00584
Tetrahydromethanopterin:CoM 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
hdrD K08264
Coenzyme M + Coenzyme B


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


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

285


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

286
287

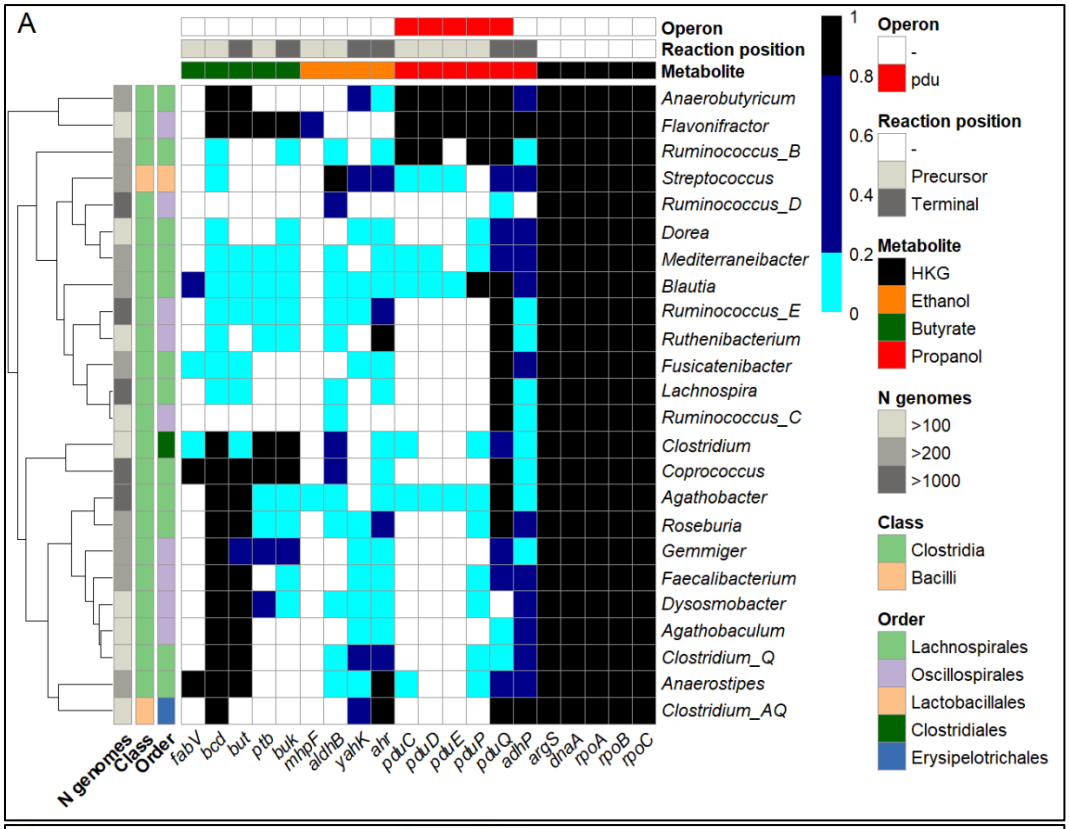
288 **Figure 4. Abundance of methane and TMA-producing genes in patient cohorts. (A)** Characterization of the terminal
289 reactions involved in the formation of microbial methane. Enzymes, EC numbers, coding genes and KOs, as well as the
290 reaction type, are indicated for each metabolic step according to the annexed legend in Fig. 2A. **(B)** Abundance of genes
291 coding for methyl-coM reductase I (*mcr*) and methyl-CoM reductase II (*mrt*). **(C)** Abundance of genes coding for CoB-CoM
292 heterodisulfide reductase (*hdr*) and tetrahydromethanopterin S-methyltransferase (*mtr*). **(D)** Abundance of *tor* operons.
293 Facets represent genic operons. Boxplot organization and statistical tests were elaborated analogously as Fig. 2C.

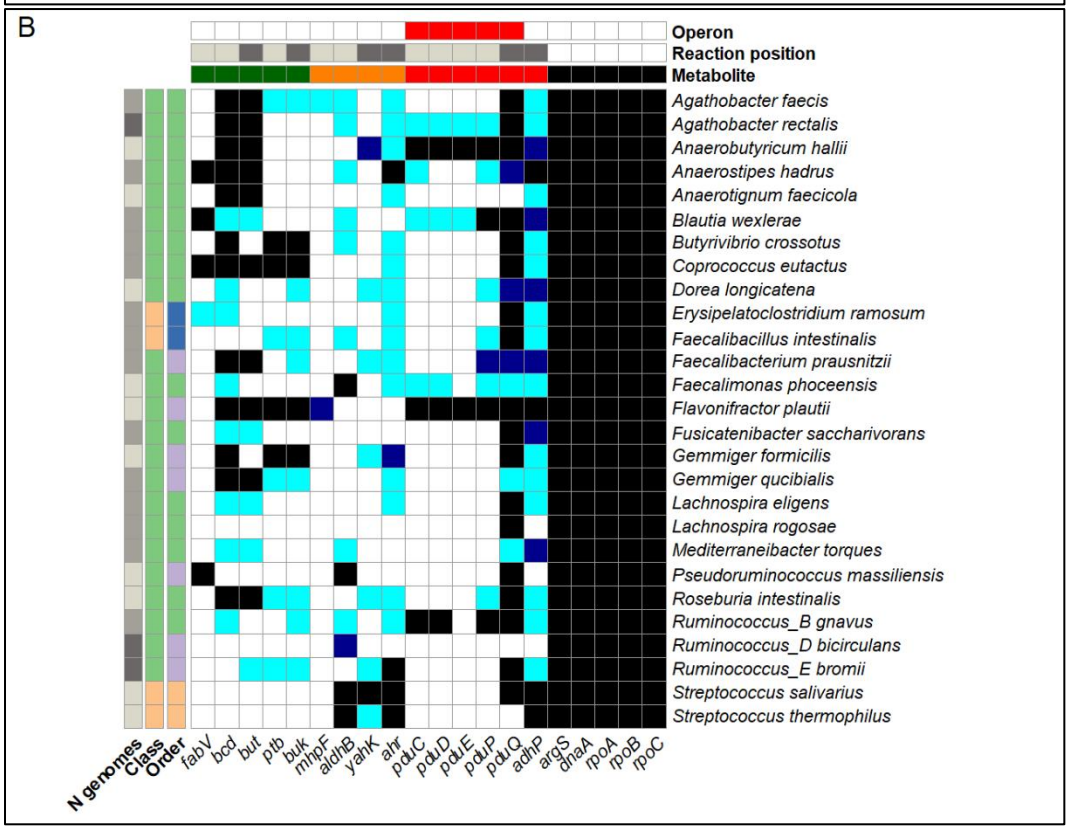
294 **Candidate metabolic genes are accessory in the GM**

295 Although this analysis identifies *A. rectalis* and multiple metabolic genes involved in the
296 formation of butyrate, SCAs, TMAO and methane as consistent taxonomic and functional signatures
297 associated with MASLD, these markers alone are not sufficiently robust to establish a causal link
298 between GM composition and MASLD phenotype. GM bacterial species have pleomorphic genomes,
299 consisting of up to thousands of genomically distinct strains that share a conserved genetic core but
300 possess an accessory genome that is highly variable among them[81,82]. This accessory genome is often
301 encoded in plasmids and other MGEs which are subject to frequent changes and can readily transfer
302 between different bacterial strains[16,83].

303 To test this hypothesis, we assessed whether candidate metabolic genes involved in MASLD
304 were core or accessory across some of the most abundant human GM genomes. Selection of genomes
305 and genic presence were determined as described in Materials and Methods. Briefly, we defined the
306 presence of a target gene in a clade as the percentage of strains encoding it, classifying it as core if
307 present in >80% of congeneric or conspecific genomes, as accessory if present in 20-80%, and as highly
308 accessory if present in <20% genomes. Most of the selected GM genera (n=24) belong to class Clostridia,
309 specifically to orders *Lachnospirales* and *Oscillospirales* (Fig. 5A). Butyrate-producing genes from the
310 butyryl-CoA pathway (*bcd* and *but*) were core in nearly half of these genera, while those from the
311 butyryl-P pathway (*ptb* and *buk*) were core only in *Flavonifractor, Clostridium* and *Coprococcus.*
312 Notably, butyrate-producing genes were highly accessory across several genera, including *Lachnospira,*

313 *Ruminococcus_B* and *E, Ruthenibacterium, Blautia*, *Fusicatenibacter*, *Dorea,* and *Mediterraneibacter*,
314 and were completely absent in *Ruminococcus_C* and *D*. In contrast, acetaldehyde and ethanol-
315 producing genes were highly accessory across all genera except *Streptococcus*, an heterolactic genus
316 that constitutively encoded *aldhB* in all its genomes (specifically in *S. epidermis, S. salivarius* and *S.*
317 *thermophilus,* Fig. 5B) and exhibited moderated presence (20-80%) of *yahK*/*ahr*. The acetaldehyde-
318 producing gene *mhpF* was nearly absent in the GM, except for a few genomes of *Flavonifractor* and
319 *Agathobacter*. Propanaldehyde and propanol-producing genes from the *pdu* operon were core in three
320 genera: *Ruminococcus_B* (*R_B. gnavus*), *Anaerobutyricum* (*A. hallii*) and *Flavonifractor* (*F. plautii*),
321 though *pduE* was unexpectedly absent in *Ruminococcus_B*. These genes were highly accessory in
322 *Blautia* (*B. wexlerae*), *Streptococcus*, *Medierraneibacter* or *Agathobacter* (*A. rectalis*). Notably, *A. hallii*
323 and *F. plautii* were the only species encoding both core butyrate and propanol-producing genes.
324 Beyond their accessory nature within GM pangenomes, candidate metabolic genes were also
325 encoded in plasmids (Fig. 5C). Genes involved in ethanol formation such as *aldhB, yahK* and *ahr* were
326 among the most widely distributed, being present in hundreds of different plasmids. In contrast,
327 butyrate-producing genes (*fabV, but, ptb,* and *buk*) and propanol-related genes (*pduC, pduD, pduE,*
328 *pduP,* and *pduQ*) were detected in fewer plasmids, further supporting their variable presence across
329 GM genomes. A notable exception was *bcd*, which exhibited a broader presence across both
330 chromosomal and plasmid-encoded sequences. These findings reinforce the idea that non-essential
331 metabolic functions, especially ethanol production, are encoded in MGEs, contributing to their
332 accessory nature within the GM.


A
Operon
Reaction position
Metabolite
Anaerobutyricum
Flavonifractor
Ruminococcus_B
Streptococcus
Ruminococcus_D
Dorea
Mediterraneibacter
Blautia
Ruminococcus_E
Ruthenibacterium
Fusicatenibacter
Lachnospira
Ruminococcus_C
Clostridium
Coprococcus
Agathobacter
Roseburia
Gemmiger
Faecalibacterium
Dysosmobacter
Agathobaculum
Clostridium_Q
Anaerostipes
Clostridium_AQ
N genomes
Class
Order
fabV
bcd
but
ptb
buk
mhpF
aldhB
yahK
ahr
pduC
pduD
pduE
pduP
pduQ
adhP
argS
dnaA
rpoA
rpoB
rpoC
1
0.8
0.6
0.4
0.2
0
Operon
-
pdu
Reaction position
-
Precursor
Terminal
Metabolite
HKG
Ethanol
Butyrate
Propanol
N genomes
>100
>200
>1000
Class
Clostridia
Bacilli
Order
Lachnospirales
Oscillospirales
Lactobacillales
Clostridiales
Erysipelotrichales


B
Operon
Reaction position
Metabolite
Agathobacter faecis
Agathobacter rectalis
Anaerobutyricum hallii
Anaerostipes hadrus
Anaerotignum faecicola
Blautia wexlerae
Butyrivibrio crossotus
Coprococcus eutactus
Dorea longicatena
Erysipelatoclostridium ramosum
Faecalibacillus intestinalis
Faecalibacterium prausnitzii
Faecalimonas phoceensis
Flavonifractor plautii
Fusicatenibacter saccharivorans
Gemmiger formicilis
Gemmiger qucibialis
Lachnospira eligens
Lachnospira rogosae
Mediterraneibacter torques
Pseudoruminococcus massiliensis
Roseburia intestinalis
Ruminococcus_B gnavus
Ruminococcus_D bicirculans
Ruminococcus_E bromii
Streptococcus salivarius
Streptococcus thermophilus
N genomes
Class
Order
fabV
bcd
but
ptb
buk
mhpF
aldhB
yahK
ahr
pduC
pduD
pduE
pduP
pduQ
adhP
argS
dnaA
rpoA
rpoB
rpoC

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

Figure 5. Presence profile of candidate metabolic genes in the human GM. (A-B) Heatmap showing the presence of candidate genes involved in MASLD across some of the most abundant clades in the human GM. The x-axis represents genes associated with butyrate and SCAs production, as well as USCGs. The y-axis lists the bacterial clades inspected. Genic presence is expressed as the percentage of **(A)** congeneric and **(B)** conspecific 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 clade. Hierarchical clustering was applied to the y-axis to group genera (top) based on genic presence similarity. **(C)** 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 influenced by multiple contributing factors, one of which is the GM. Numerous studies have tried to ascribe a causality between GM composition and MASLD. However, taxonomic signatures in key GM clades, such as *Blautia* and *Roseburia* spp., have been inconsistently associated with MASLD progression, with studies reporting both positive and negative correlations[23]. These issues are normally linked to the use of 16S rRNA-based profiling in metataxonomic studies, which often lacks resolution at species and strain level[84]. To address this limitation and assess if there is indeed a specific clade robustly associated with MASLD, we utilized multiple marker genes for taxonomic classification, in an attempt to provide a more comprehensive view of microbial abundances

354 in health and disease states. Using MetaPhlan, we profiled the taxonomic composition of the GM in
355 metagenomic fecal samples of healthy and MASLD donors across three independent cohorts. We
356 identified 28 bacterial genera shared across all cohorts (Fig. 1A), collectively accounting for 30-60% of
357 the total GM abundance (Fig. 1B). Among these, five candidate genera -*Gemmiger*, *Streptococcus*,
358 *Mediterraneibacter*, *Ruminococcus* and *Agathobacter* (Fig. 1C)- and two species, *Agathobacter rectalis*
359 and *Bacteroides uniformis* (Fig. 1E), exhibited consistent differences in their relative abundance
360 between groups across all the cohorts. These taxa rank among the most abundant clades in the GM
361 (Fig. 1C, Supplementary Fig. 1). However, only *Agathobacter rectalis* showed statistically significant
362 differences. Given its high abundance in the GM, its coherent abundance trend at both genus and
363 species level, and its consistent depletion in MASLD, *A. rectalis* emerges as a robust taxonomic marker
364 for the disease.

365 Although this observation has been previously reported[18,21], this result should be interpreted
366 with caution due to ongoing debates in bacterial nomenclature, as *A. rectalis* is a newer designation
367 that has replaced *Eubacterium rectale*[85]. Updates in taxonomic assignments pose challenges, as newer
368 species or genera may not be consistently reflected across different databases or studies[86,87]. Also,
369 taxonomic classifications themselves, whether based solely on 16S rRNA or on multiple markers, limit
370 our ability to reliably infer GM changes. Bacterial genome plasticity, driven by MGEs, can blur the lines
371 of taxonomy-phenotype associations, as they contribute to a vast strain-level diversity[81] that remains
372 undetected by taxonomic profiling methods. In fact, we have previously reported that geno-metabolic
373 shifts in GM propionate production associated with inflammatory bowel disease do not align with
374 species-level variations[88].

375 GM associated metabolism contributes to MASLD by disrupting the integrity of the GLA. For this
376 reason, we focused on identifying abundance differences in the genes responsible for producing
377 metabolites potentially implicated in MASLD across several patient cohorts, i.e. butyrate, SCAs, TMA
378 and methane. We isolated the GM gene families encoding the enzymes involved in the biosynthesis of
379 these metabolites and aligned them against the metagenomic samples to quantify their abundance.
380 We found that butyrate-producing genes are depleted in the GM of MASLD patients (Fig. 2B),
381 particularly those involved in the crotonyl-butyryl CoA axis: *fabV* (butyryl-CoA oxidoreductase), *bcd*
382 (butyryl-CoA dehydrogenase) and *but* (butyryl-CoA:acetate-CoA transferase) (Fig. 2A). While *ptb* and
383 *buk* were also higher in both healthy subcohorts, their levels did not differ significantly between initial
384 and advanced MASLD groups within Cohort 2. This suggests that CoA transfer is the predominant
385 pathway for butyrate production in the GM, as previously reported[89]. The smaller differences in *ptb*
386 and *buk* in Cohort 2 may reflect a common MASLD phenotype across different fibrosis stages, along
387 with the limited sample size of this cohort compared to Cohort 3. Additionally, two other enzymes -
388 structured as two-subunit systems-, can transfer CoA from different cofactors to butyrate to produce
389 butyryl-CoA: butyryl-CoA:acetoacetate-CoA transferase (encoded by *atoD* and *atoA*) and acetate CoA-
390 transferase (encoded by *ctfA* and *ctfB*). These were not included in this study as they catalyze the
391 reverse reaction, consuming butyrate to form butyryl-CoA[90]. Notably, the depletion of butyrate genes
392 was not observed in MASLD patients from Cohort 1.

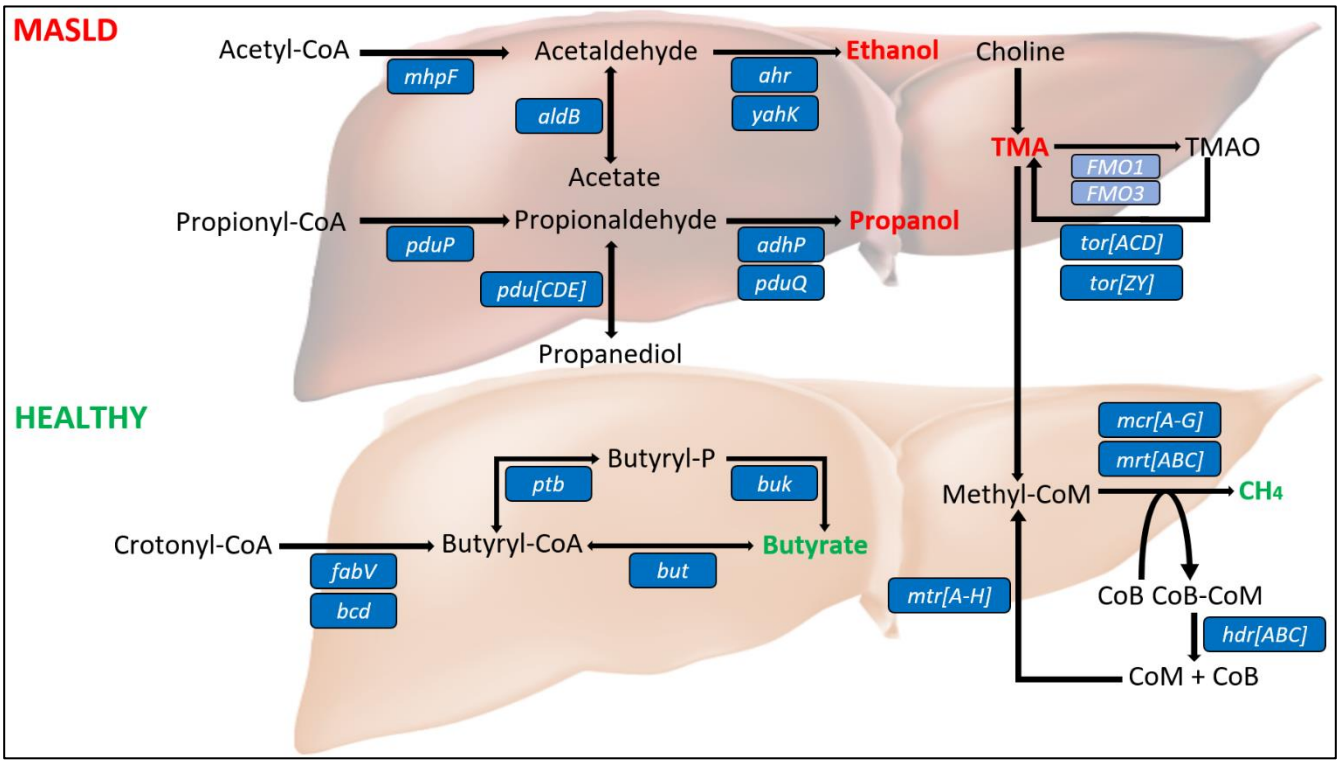
393 Contrarily, genes involved in the formation of ethanol and propanol through connected
394 fermentative reactions are increased in MASLD (Fig. 3). Genes *mhpF* and *aldB* -encoding two aldehyde
395 dehydrogenases that produce ethanol-, and *yahK* and *ahr* -encoding two NADPH-dependent aldehyde
396 reductases that form acetaldehyde- are significantly increased in the GM of MASLD patients compared

397 to the healthy groups (Fig. 3B). Genes involved in the production of 1-propanol structured in operon
398 *pdu* are also increased in MASLD (Fig. 3C). Specifically, *pduCDE* -encoding a propanediol dehydratase-,
399 *pduP* -encoding a propionaldehyde dehydrogenase- and *pduQ* and *adhP* -encoding two distinct
400 propanol dehydrogenases-. The marked differential abundances of propanol-producing genes,
401 organized in the *pdu* operon and *adhP*, which have been largely unexplored in the context of liver
402 disease, suggests that 1-propanol produced in bacterial microcompartments may play an even more
403 critical role than ethanol in MASLD onset and progression.

404 The abundance of genes in the TMA-methane metabolic axis is also disrupted in MASLD. Genes
405 involved in methane formation (Fig. 4A) are depleted in the pathology, including both MCR I and II
406 coding-genes (Fig. 4B), which reduce methyl-CoM to methane and *mtr*, *hdr* (Fig. 4C), and Mtb / Mta-
407 coding genes (Supplementary Fig. 2), dedicated to regenerating methyl-CoM. In contrast, TMAO-
408 reductive genes, organized in *torCAD* and *torZY* operons and responsible for regenerating TMA, are
409 increased in MASLD (Fig. 3D). These results suggest that methane production is impaired in MASLD. The
410 opposing trend in the abundance of *tor* operons further indicates that this impairment stems from a
411 metabolic bottleneck that leads to TMA accumulation instead of its conversion to methyl-CoM, effect
412 that is driven by TMAO reductases I and II.

413 Taken together, these results reveal a geno-metabolic alteration associated with MASLD,
414 characterized by an increase in genes involved in the production of SCAs and TMA, and a concurrent
415 depletion of genes responsible for butyrate and methane production (Fig. 6). We confirmed that genic
416 abundance differences were driven by the MASLD phenotype by showing no differences in USCG
417 abundance between groups (Supplementary Fig. 3). These findings support the hypothesis that the
418 accumulation of endogenous alcohols may represent a key metabolic feature of MASLD, occurring

419 alongside the displacement of butyrate-producing taxa, as previously proposed[33]. Moreover, the
420 observed decrease in *A. rectalis* abundance partially explain the reduction in butyrate levels, as this
421 organism is a known GM butyrate producer[85].



422

423 **Figure 6. Geno-metabolic scheme involved in MASLD.** Proposed scheme showing the GM genic alterations related to
424 butyrate, SCAs, TMA and methane production in MASLD. Ethanol, propanol and TMA-producing genes are increased in the
425 pathology (top), whereas butyrate and methane-producing genes are decreased (bottom). Blue boxes indicate the catalyzing
426 genes whose abundance is altered in each metabolic step.

427 Importantly, we found that metabolic genes associated with MASLD exhibit a highly accessory
428 profile within the GM (Fig. 5). Butyrate-producing genes are highly accessory in nearly half of the most
429 abundant GM genera, while SCA-producing genes display an even more pronounced accessory pattern.
430 Among the two main butyrate biosynthesis routes, genes from the butyryl-CoA pathway are more
431 pervasive in the GM than those from the butyryl-P pathway. Notably, *Phocaeicola* and *Bacteroides* -the
432 two most abundant GM genera- were excluded from this analysis due to a limited number of high-
433 quality genomes passing our completeness filters, despite their extensive genomic representation in

434 the UHGG. Interestingly, the two key genes in the butyryl-CoA pathway (*bcd* and *but*) are core in *A.*
435 *rectalis*. Nonetheless, some strains of this known butyrate-producing species also harbor genes for
436 ethanol and propanol production, which our findings suggest may contribute to MASLD onset. This
437 highlights a critical point: taxonomic signatures alone are insufficient for associating bacterial clades
438 with disease. Genome plasticity introduces substantial strain-level metabolic variability in the GM,
439 leading to distinct phenotypic traits even within the same species: an effect well-documented in both
440 pathogenic and commensal *E. coli* strains[91,92].

441 Further supporting their mobility, candidate genes are frequently encoded in plasmids (Fig. 5C)
442 with ethanol formation genes (*aldhB*, *yahK*, and *ahr*) being the most widely distributed, while butyrate
443 and propanol-related genes are present to a lesser extent. Since non-essential, adaptive metabolic
444 pathways are often encoded in plasmids across many bacterial species[92–94], their presence likely drives
445 phenotypic differences in MASLD without altering the overall species composition detected by
446 conventional metataxonomic marker genes. Given that MASLD is partially driven by bacterial
447 metabolism, strains from different species probably may produce similar liver-impacting metabolites,
448 potentially contributing to MASLD pathogenesis.

449 **4. Conclusions**

450 Our study reveals that MASLD is marked by distinct geno-metabolic alterations in the GM that
451 are not captured by conventional taxonomic profiling. Specifically, we observed an increase in genes
452 involved in the production of SCAs and TMA, coupled with a depletion of genes responsible for butyrate
453 and methane production. These results confirm multiple previous findings and suggest a broader
454 metabolic shift that may contribute to the pathogenesis of MASLD, with the displacement of beneficial
455 butyrate-producing taxa such as *Agathobacter rectalis*. Importantly, our work underscores the value of

analyzing metagenomes at gene level to extract additional phenotypic signatures beyond what 16S rRNA or marker gene-based approaches can offer. By integrating metabolic pathway analysis with the isolation of specific gene families, and by examining gene abundance, accessory gene profiles and plasmid-mediated gene mobilization, we have significantly advanced our understanding of bacterial metabolism in MASLD. This comprehensive approach not only enhances our vision of the microbial etiology of MASLD, but also paves the way for improved bacterial analysis in clinical settings, offering new avenues for diagnosis and therapeutic intervention in liver pathologies.

Materials and Methods

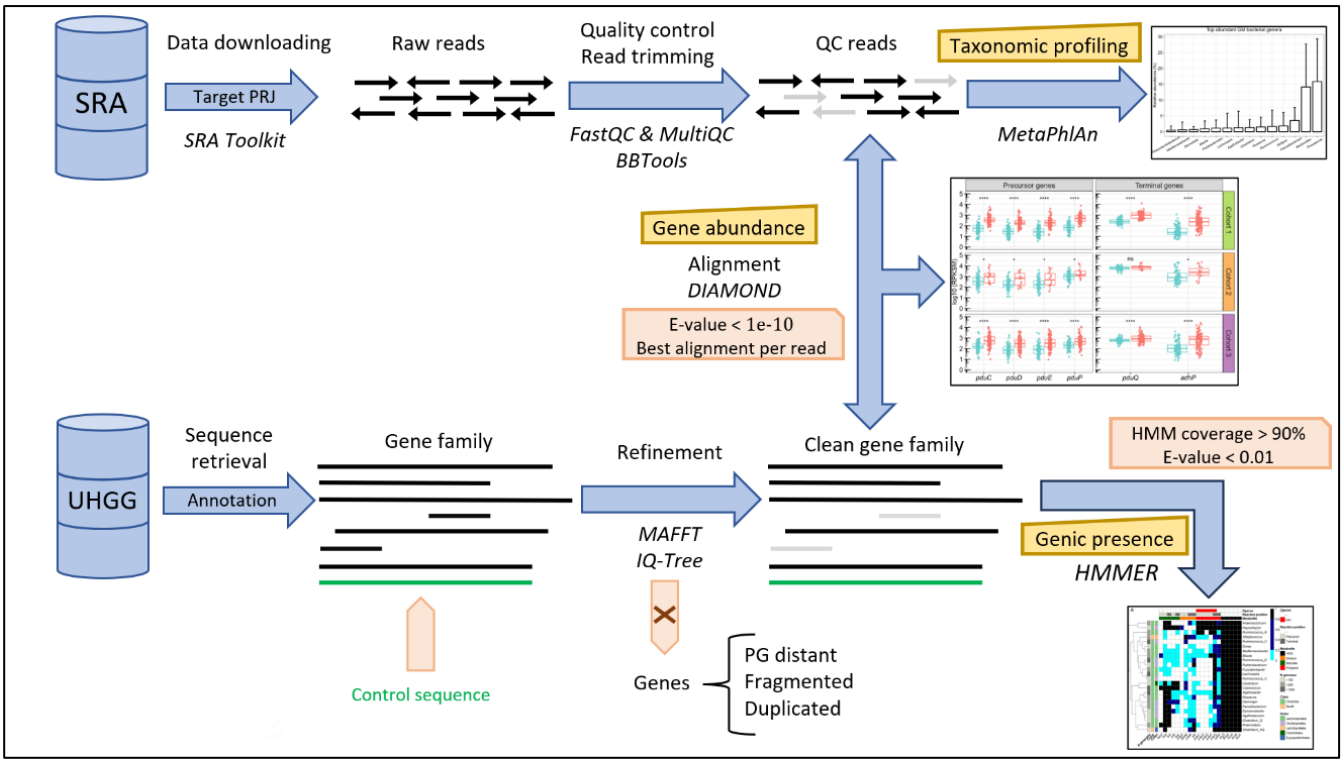


Figure 7. Computational workflow used in this study. This diagram summarizes the full computational methodology for processing whole paired-end metagenomic samples, including data retrieval from the SRA, read quality control and taxonomic profiling. Target gene families were isolated and their abundance quantified by alignment against the sequencing libraries. Additionally, the genic presence of target genes was examined across GM genomes and annotated plasmids was inspected. Blue arrows indicate the progression between workflow steps, with the corresponding software shown in black beneath each arrow. Filtering steps are highlighted in pink boxes.

471 ***Downloading of sequencing metagenomic data***

472 A total of 554 metagenomic libraries generated from the fecal samples of MASLD and healthy
473 subjects were downloaded from the SRA using the accession numbers corresponding to five different
474 studies with SRA Toolkit´s fastq-dump[95]. These samples were retrieved using the accession numbers
475 corresponding to five different studies, and they were stratified in three MASLD cohorts, as detailed in
476 Table 1. and they were stratified in three MASLD cohorts as detailed in Table 1. Information on the
477 collection and processing of fecal samples, bacterial DNA extraction and sequencing protocols can be
478 found in the original studies. Additionally, we retrieved the medical metadata associated with each
479 sample donor to confirm their MASLD status and classify samples into two phenotypic groups of
480 comparison within each cohort. Samples lacking a clearly defined MASLD status were excluded from
481 this study. All data was downloaded between March 2022 and March 2024

Cohort	Phenotypic group	N samples	Reference	Accession number
1	Healthy control	117	96	PRJNA544527
	MASLD, initial fibrosis (F0-F2)	73	97	PRJEB14215
		48	98	PRJNA420817
2	MASLD, initial fibrosis (F0-F2)	72	18	PRJNA373901
	MASLD, advanced fibrosis (F3-F4)	14		
3	Healthy control	105	99	PRJEB6337
	Cirrhosis MASLD	125		

482 **Table 1: MASLD patient cohorts analyzed in this study.**

483 Each cohort (column 1) was divided into two phenotypic comparison groups: control vs. disease (column 2). The cohorts
484 include *N* MASLD patients and corresponding whole-metagenomic stool samples derived from them (column 3). The disease
485 group in Cohort 1 consists of 73 and 48 MASLD patients with initial fibrosis from[97,98]. Cohort 2 was divided into initial fibrosis
486 (control) and advanced fibrosis. Patients were originally recruited in the listed studies (column 4), with metagenomic data
487 and medical metadata downloaded from the indicated BioProject accession numbers (column 5).

488

489 ***Description of MASLD patient cohorts***

490 Cohort 1 consists of two sets of European, obese MASLD patients aged 20-64 years with initial
491 fibrosis (F0-F2) and varying degrees of steatosis: 73 patients from [97] and 10 patients from [98]. In the latter
492 group, 48 fecal samples were collected at five time points during the study (two samples from day 1
493 were not obtained). Diagnoses were confirmed through hepatic biopsy performed during obesity
494 surgery, and none of the subjects had alcoholic liver disease, hepatitis or diabetes. The control group
495 for comparison includes 117 healthy patients from [96]. Cohort 2 comprises 86 American MASLD patients
496 from [18], including 72 with initial (F0-F2) and 14 with advanced fibrosis (F3-F4). Diagnoses were
497 confirmed through hepatic biopsy and magnetic resonance imaging, and none of the patients had
498 alcoholic liver disease, hepatitis or diabetes. Note that in this study, the initial fibrosis group is referred
499 to as "healthy" for group comparison purposes, despite including patients with early-stage MASLD.
500 Cohort 3 includes 230 Han Chinese patients from [99]: 125 diagnosed with cirrhotic MASLD and 105 with
501 no hepatic pathology. None of these subjects had hepatitis. All cohorts consist of one metagenomic
502 stool sample per patient, except for the 48 samples collected at multiple time points from 10 patients
503 in [98].

504

505 ***Quality control of metagenomic reads***

506 Sequencing reads were trimmed to remove Illumina adapter remnants and low-quality regions
507 (<Q25) using BBDuk from the BBTools suite (version 37.62)[100]. Read quality was assessed with FastQC
508 (version 0.12.1)[101] and summarized using MultiQC (version 1.22.3)[102].

509

510

511 ***Taxonomic signatures associated with MASLD***

512 To determine whether any bacterial clade is systematically increased or decreased in MASLD
513 across the three cohorts, we profiled the taxonomic composition of GM communities in fecal samples
514 using MetaPhlAn (version 4.0.6)[24]. Briefly, MetaPhlAn estimates the relative abundance of bacterial
515 taxa by aligning the metagenomic reads against a curated database of 5 million unique clade-specific
516 marker genes derived from 1 million microbial genomes spanning 26,970 species-level genome bins.
517 These marker genes include a range of taxonomically informative sequences, such as those from rRNA
518 genes (16S, 23S), RNA polymerase subunit genes (*rpoB*), DNA gyrase subunits (*gyrB*), and chaperone
519 proteins (*dnaK*), which are present in most prokaryotic genomes. The diversity of these markers allows
520 MetaPhlAn to provide highly accurate taxonomic classification across a wide spectrum of bacterial
521 clades, from the species level to broader taxonomic ranks, without being limited to 16S rRNA alone.

522 For this study, we used the database version *mpa_vOct22_CHOCOPhlAnSGB_202212*,
523 downloaded on August 22, 2023, and added the *--ignore_eukaryotes* flag to skip profiling eukaryotic
524 organisms. Post-processing of the MetaPhlAn output involved a) removing low-abundant clades (i.e.,
525 those with an average relative abundance of 0 in at least one cohort), b) excluding taxa predicted as
526 "unclassified", and c) verifying that excluded clades were not solely present in only one of the
527 comparison groups, in order to avoid what could be very clear taxonomic markers. To identify the most
528 abundant clades present in the GM, we selected genera with an average relative abundance above
529 0.5% (Fig. 1C) and species above 0.2% (Fig. 1E). Notedly, in this step we faced an issue regarding the
530 status of taxonomic assignments, as several species required manual updates to their nomenclature
531 assigned by MetaPhlAn. Specifically, *Eubacterium rectale* was reassigned to *Agathobacter rectalis*[85],
532 *Ruminococcus torques* to *Mediterraneibacter torques*[103] and *Roseburia faecis* to *Agathobacter faecis*[104].

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

534 We constructed families of the human GM-metabolic genes encoding the enzymes involved in the
535 production of butyrate, ethanol, propanol, TMA, methane and their precursor metabolites, as
536 represented in Figs. 2A, 3A and 4A. The Unified Human Gastrointestinal Genome (UHGG, version
537 2.0.2)[81] served as the reference database for this purpose. This resource, comprising 289,232 bacterial
538 and archaeal genomes and more than 665 million proteins, represents the most comprehensive
539 collection of characterized human GM genes to date. Each gene family was built by retrieving a seed of
540 sequences based on gene and protein annotation from the UHGG, assessing their similarity with lab-
541 validated enzymatic sequences when available. Validated gene sequences were retrieved from
542 MetaCyc[105] and used as controls to assess their phylogenetic proximity to the retrieved gene seeds.
543 Gene families were aligned using MAFFT (version 7.271)[106] and phylogenetic trees were constructed
544 with IQ-TREE (version 2.0.3)[107]. Low phylogenetic distances between control and retrieved sequences
545 served as a criterion for defining gene families. Duplicated sequences, gene fragments and
546 phylogenetically distant genes were excluded after inspecting the trees. To improve gene identification
547 and classification, we then built Hidden Markov Model (HMM) profiles for each target gene family using
548 *HMMER hmpress* (version 3.4)[108]. MetaCyc[105] and KEGG[109] were also used to validate gene and enzyme
549 names, KEGG Orthology (KOs) groups, the biochemical reactions involved in the formation of each
550 metabolite, and their enzyme commission (EC) numbers. The same protocol was applied to isolate five
551 USCGs, used as controls. These USCGs and their encoded proteins were: *argS* (arginyl-tRNA synthetase),
552 *dnaA* (chromosomal replication initiator protein DnaA), *rpoA*, *rpoB* and *rpoC* (alpha, beta and gamma
553 subunits of the DNA-directed RNA polymerase, respectively).

554

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

556 We quantified the abundance of target genes in the metagenomic samples from the three cohorts
557 by aligning the quality-controlled reads against the isolated gene families using DIAMOND (version
558 2.0.14)[110]. Alignments were filtered to retain only the best match for each read and gene family
559 sequence by using the *"--max-hsps 1"* option, with an E-value threshold <1E-10. These parameters were
560 based on experimentally validated recommendations for detecting genes in fecal metagenomic
561 samples with DIAMOND[111]. The total number of aligned reads for each gene family was normalized
562 using a modified RPKM formula, referred to as RPKSM (reads per kilobase per size per million reads).
563 This measure was used as an indicator of gene abundance (Figs. 2B, 3B-C and 4B-D; and Supplementary
564 Figs. 1 and 2), and it is defined as the number of aligned reads per kilobase per gene family size per
565 million reads, to account for variations in gene length, gene family size (i.e., the number of sequences
566 within each gene family) and library size, respectively.

567

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

569 To determine whether candidate metabolic genes involved in MASLD were core or accessory across
570 the human GM, we analyzed their presence in both UHGG genomes and curated RefSeq200 plasmids[112].
571 Curation consisted in eliminating sequences that corresponded to partial plasmid DNA sequences,
572 unassignable hosts, or PacBio internal control sequences, as detailed previously[16]. Since many UHGG
573 genomes originate from metagenome-assembled or single-cell-amplified genomes, they may be
574 incomplete, potentially leading to false positive annotations. To minimize this risk, we included only
575 UHGG genomes with >95% completeness and containing the five USCGs described above. ORFs were
576 predicted using Prodigal (version 2.6.3)[113], retaining only those >90% of the HMM size. We then used

577 HMMER *hmmscan* with the following parameters: protein-pHMM coverage threshold >90%, E-value
578 <0.01 and independent E-value (i-E value) <0.01, strictly enforcing the 90% threshold to avoid capturing
579 ORFs with similar protein domains. This approach yielded 31,227 genomes classified as complete and
580 containing the five USCGs.

581 To classify genes as core or accessory, we defined their presence as the percentage of strains within
582 a given genus and species that encode the gene, considering only the most abundant clades in the GM
583 with >100 genomes meeting the completeness criteria. Genes were categorized as core if present in
584 >80% of congeneric (Fig. 5A) or conspecific (Fig. 5B) genomes, as accessory if present in 20-80%, and as
585 highly accessory if present in <20% genomes. Additionally, we assessed whether these genes could be
586 encoded in annotated plasmids. Using the same constraints, we predicted ORFs from the RefSeq200
587 plasmid collection, which includes 23,309 plasmids encoding over 2 million ORFs. ORFS were retained
588 if they met the >90% HMM size threshold, with *hmmscan* parameters identical to those applied in the
589 UHGG genome analysis (Fig. 5C).

590

591 ***R packages and statistical analysis***

592 Data manipulation was performed using R (version 4.1.3) and the *tidyverse* package (version
593 1.3.2)[114]. Figures were generated with the R packages *ggplot2* (version 3.4.2)[115], *ggpubr* (version
594 0.6.0)[116] and *pheatmap* (version 1.0.12)[117]. Mann-Whitney tests with a Benjamini-Hochberg adjustment
595 for multiple testing were applied on the relative abundance of bacterial clades (Figs. 1D and 1E) and
596 genic abundance (Figs. 2B, 3B-C, 4B-D, and Supplementary Figs. 2 and 3) between phenotypic groups
597 in the metagenomic fecal samples of each cohort to assess the significance of the differences. The

598 statistical analysis and its graphic representation was conducted using the R package *rstatix* (version
599 0.7.2)[118].

600

601 **References**

602 1. Riazi, K. *et al.* The prevalence and incidence of NAFLD worldwide: a systematic review and meta-analysis.
603 *Lancet Gastroenterol. Hepatol.* **7**, 851–861 (2022).
604 2. EASL–EASD–EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. *J.*
605 *Hepatol.* **64**, 1388–1402 (2016).
606 3. Allen, A. M. *et al.* Clinical course of non-alcoholic fatty liver disease and the implications for clinical trial
607 design. *J. Hepatol.* **77**, 1237–1245 (2022).
608 4. Powell, E. E., Wong, V. W.-S. & Rinella, M. Non-alcoholic fatty liver disease. *The Lancet* **397**, 2212–2224
609 (2021).
610 5. Yki-Järvinen, H. Non-alcoholic fatty liver disease as a cause and a consequence of metabolic syndrome.
611 *Lancet Diabetes Endocrinol.* **2**, 901–910 (2014).
612 6. Martín-Mateos, R. & Albillos, A. The Role of the Gut-Liver Axis in Metabolic Dysfunction-Associated Fatty
613 Liver Disease. *Front. Immunol.* **12**, 660179 (2021).
614 7. Carpino, G. *et al.* Increased Liver Localization of Lipopolysaccharides in Human and Experimental NAFLD.
615 *Hepatology* **72**, 470–485 (2020).
616 8. Li, Q., Rempel, J. D., Yang, J. & Minuk, G. Y. The Effects of Pathogen-Associated Molecular Patterns on
617 Peripheral Blood Monocytes in Patients with Non-alcoholic Fatty Liver Disease. *J. Clin. Exp. Hepatol.* **12**,
618 808–817 (2022).
619 9. Miele, L. *et al.* Increased intestinal permeability and tight junction alterations in nonalcoholic fatty liver
620 disease†. *Hepatology* **49**, 1877–1887 (2009).

621 10. Wang, W. *et al.* Tauroursodeoxycholic acid inhibits intestinal inflammation and barrier disruption in mice
622 with non-alcoholic fatty liver disease. *Br. J. Pharmacol.* **175**, 469–484 (2018).
623 11. Kuang, J. *et al.* Hyodeoxycholic acid alleviates non-alcoholic fatty liver disease through modulating the gut-
624 liver axis. *Cell Metab.* **35**, 1752-1766.e8 (2023).
625 12. Yang, C. *et al.* Altered gut microbial profile accompanied by abnormal short chain fatty acid metabolism
626 exacerbates nonalcoholic fatty liver disease progression. *Sci. Rep.* **14**, 22385 (2024).
627 13. Zhang, R. *et al.* Gut microbial metabolites in MASLD: Implications of mitochondrial dysfunction in the
628 pathogenesis and treatment. *Hepatol. Commun.* **8**, (2024).
629 14. Del Chierico, F. *et al.* Gut microbiota profiling of pediatric nonalcoholic fatty liver disease and obese
630 patients unveiled by an integrated meta-omics-based approach. *Hepatology* **65**, 451–464 (2017).
631 15. Lee, G. *et al.* Distinct signatures of gut microbiome and metabolites associated with significant fibrosis in
632 non-obese NAFLD. *Nat. Commun.* **11**, 4982 (2020).
633 16. Redondo-Salvo, S. *et al.* Pathways for horizontal gene transfer in bacteria revealed by a global map of their
634 plasmids. *Nat. Commun.* **11**, 3602 (2020).
635 17. Lanthier, N. *et al.* Microbiota analysis and transient elastography reveal new extra-hepatic components of
636 liver steatosis and fibrosis in obese patients. *Sci. Rep.* **11**, 659 (2021).
637 18. Loomba, R. *et al.* Gut Microbiome-Based Metagenomic Signature for Non-invasive Detection of Advanced
638 Fibrosis in Human Nonalcoholic Fatty Liver Disease. *Cell Metab.* **25**, 1054-1062.e5 (2017).
639 19. Boursier, J. *et al.* The severity of nonalcoholic fatty liver disease is associated with gut dysbiosis and shift in
640 the metabolic function of the gut microbiota. *Hepatology* **63**, 764–775 (2016).
641 20. Caussy, C. *et al.* A gut microbiome signature for cirrhosis due to nonalcoholic fatty liver disease. *Nat.*
642 *Commun.* **10**, 1406 (2019).
643 21. Oh, T. G. A Universal Gut-Microbiome-Derived Signature Predicts Cirrhosis.

644 22. Shen, F. *et al.* Gut microbiota dysbiosis in patients with non-alcoholic fatty liver disease. *Hepatobiliary*
645 *Pancreat. Dis. Int.* **16**, 375–381 (2017).
646 23. Aron-Wisnewsky, J. *et al.* Gut microbiota and human NAFLD: disentangling microbial signatures from
647 metabolic disorders. *Nat. Rev. Gastroenterol. Hepatol.* **17**, 279–297 (2020).
648 24. Blanco-Míguez, A. *et al.* Extending and improving metagenomic taxonomic profiling with uncharacterized
649 species using MetaPhlAn 4. *Nat. Biotechnol.* **41**, 1633–1644 (2023).
650 25. McPherson, S. *et al.* Age as a Confounding Factor for the Accurate Non-Invasive Diagnosis of Advanced
651 NAFLD Fibrosis. *Am. J. Gastroenterol.* **112**, 740–751 (2017).
652 26. Ponziani, F. R., Gasbarrini, A. & Pompili, M. NAFLD or comorbidities, that is the question. *J. Hepatol.* **73**, 723
653 (2020).
654 27. Chu, H., Duan, Y., Yang, L. & Schnabl, B. Small metabolites, possible big changes: a microbiota-centered
655 view of non-alcoholic fatty liver disease. *Gut* **68**, 359–370 (2019).
656 28. Dai, X. *et al.* Microbial Metabolites: Critical Regulators in NAFLD. *Front. Microbiol.* **11**, 567654 (2020).
657 29. Martin-Grau, M. & Monleón, D. The Role of Microbiota-Related Co-Metabolites in MASLD Progression: A
658 Narrative Review. *Curr. Issues Mol. Biol.* **46**, 6377–6389 (2024).
659 30. Oliphant, K. & Allen-Vercoe, E. Macronutrient metabolism by the human gut microbiome: major
660 fermentation by-products and their impact on host health. *Microbiome* **7**, 91 (2019).
661 31. Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. Gut biogeography of the bacterial microbiota. *Nat. Rev.*
662 *Microbiol.* **14**, 20–32 (2016).
663 32. Albillos, A., De Gottardi, A. & Rescigno, M. The gut-liver axis in liver disease: Pathophysiological basis for
664 therapy. *J. Hepatol.* **72**, 558–577 (2020).
665 33. Iruzubieta, P., Medina, J. M., Fernández-López, R., Crespo, J. & De La Cruz, F. A Role for Gut Microbiome
666 Fermentative Pathways in Fatty Liver Disease Progression. *J. Clin. Med.* **9**, 1369 (2020).

667 34. Miller, T. L. & Wolin, M. J. Pathways of acetate, propionate, and butyrate formation by the human fecal
668 microbial flora. *Appl. Environ. Microbiol.* **62**, 1589–1592 (1996).
669 35. Koh, A., De Vadder, F., Kovatcheva-Datchary, P. & Bäckhed, F. From Dietary Fiber to Host Physiology: Short-
670 Chain Fatty Acids as Key Bacterial Metabolites. *Cell* **165**, 1332–1345 (2016).
671 36. Morrison, D. J. & Preston, T. Formation of short chain fatty acids by the gut microbiota and their impact on
672 human metabolism. *Gut Microbes* **7**, 189–200 (2016).
673 37. Blaak, E. E. *et al.* Short chain fatty acids in human gut and metabolic health. *Benef. Microbes* **11**, 411–455
674 (2020).
675 38. Donohoe, D. R. *et al.* The Microbiome and Butyrate Regulate Energy Metabolism and Autophagy in the
676 Mammalian Colon. *Cell Metab.* **13**, 517–526 (2011).
677 39. Wang, H.-B., Wang, P.-Y., Wang, X., Wan, Y.-L. & Liu, Y.-C. Butyrate Enhances Intestinal Epithelial Barrier
678 Function via Up-Regulation of Tight Junction Protein Claudin-1 Transcription. *Dig. Dis. Sci.* **57**, 3126–3135
679 (2012).
680 40. Gao, Z. *et al.* Butyrate Improves Insulin Sensitivity and Increases Energy Expenditure in Mice. *Diabetes* **58**,
681 1509–1517 (2009).
682 41. Li, Z. *et al.* Butyrate reduces appetite and activates brown adipose tissue via the gut-brain neural circuit.
683 *Gut* **67**, 1269–1279 (2018).
684 42. De Vadder, F. *et al.* Microbiota-Generated Metabolites Promote Metabolic Benefits via Gut-Brain Neural
685 Circuits. *Cell* **156**, 84–96 (2014).
686 43. Sanna, S. *et al.* Causal relationships among the gut microbiome, short-chain fatty acids and metabolic
687 diseases. *Nat. Genet.* **51**, 600–605 (2019).
688 44. Endo, H., Niioka, M., Kobayashi, N., Tanaka, M. & Watanabe, T. Butyrate-Producing Probiotics Reduce
689 Nonalcoholic Fatty Liver Disease Progression in Rats: New Insight into the Probiotics for the Gut-Liver Axis.
690 *PLoS ONE* **8**, e63388 (2013).

691 45. Jin, C. J., Sellmann, C., Engstler, A. J., Ziegenhardt, D. & Bergheim, I. Supplementation of sodium butyrate
692 protects mice from the development of non-alcoholic steatohepatitis (NASH). *Br. J. Nutr.* **114**, 1745–1755
693 (2015).
694 46. Zhou, D. *et al.* Sodium butyrate attenuates high-fat diet-induced steatohepatitis in mice by improving gut
695 microbiota and gastrointestinal barrier. *World J. Gastroenterol.* **23**, 60 (2017).
696 47. Zhou, D. *et al.* Sodium butyrate reduces high-fat diet-induced non-alcoholic steatohepatitis through
697 upregulation of hepatic GLP-1R expression. *Exp. Mol. Med.* **50**, 1–12 (2018).
698 48. Zhou, D. *et al.* Total fecal microbiota transplantation alleviates high-fat diet-induced steatohepatitis in mice
699 via beneficial regulation of gut microbiota. *Sci. Rep.* **7**, 1529 (2017).
700 49. Louis, P. & Flint, H. J. Formation of propionate and butyrate by the human colonic microbiota. *Environ.*
701 *Microbiol.* **19**, 29–41 (2017).
702 50. Zhu, L. *et al.* Characterization of Gut Microbiomes in Nonalcoholic Steatohepatitis (NASH) Patients: A
703 Connection Between Endogenous Alcohol and NASH. *Hepatology* **57**, 601–609 (2013).
704 51. Michail, S. *et al.* Altered gut microbial energy and metabolism in children with non-alcoholic fatty liver
705 disease. *FEMS Microbiol. Ecol.* **91**, 1–9 (2015).
706 52. Meijnikman, A. S. *et al.* Microbiome-derived ethanol in nonalcoholic fatty liver disease. *Nat. Med.* **28**,
707 2100–2106 (2022).
708 53. Yuan, J. *et al.* Fatty Liver Disease Caused by High-Alcohol-Producing Klebsiella pneumoniae. *Cell Metab.* **30**,
709 675-688.e7 (2019).
710 54. Chen, X. *et al.* Endogenous ethanol produced by intestinal bacteria induces mitochondrial dysfunction in
711 non-alcoholic fatty liver disease. *J. Gastroenterol. Hepatol.* **35**, 2009–2019 (2020).
712 55. Fromenty, B. & Roden, M. Mitochondrial alterations in fatty liver diseases. *J. Hepatol.* **78**, 415–429 (2023).
713 56. Ruuskanen, M. O. *et al.* Links between gut microbiome composition and fatty liver disease in a large
714 population sample. *Gut Microbes* **13**, 1888673 (2021).

715 57. Malaguarnera, G. Gut microbiota in alcoholic liver disease: Pathogenetic role and therapeutic perspectives.
716 *World J. Gastroenterol.* **20**, 16639 (2014).
717 58. Schut, G. J. *et al.* Tungsten enzymes play a role in detoxifying food and antimicrobial aldehydes in the
718 human gut microbiome. *Proc. Natl. Acad. Sci.* **118**, e2109008118 (2021).
719 59. Pony, P., Rapisarda, C., Terradot, L., Marza, E. & Fronzes, R. Filamentation of the bacterial bi-functional
720 alcohol/aldehyde dehydrogenase AdhE is essential for substrate channeling and enzymatic regulation. *Nat.*
721 *Commun.* **11**, 1426 (2020).
722 60. Hitschler, L., Nissen, L. S., Kuntz, M. & Basen, M. Alcohol dehydrogenases AdhE and AdhB with broad
723 substrate ranges are important enzymes for organic acid reduction in Thermoanaerobacter sp. strain X514.
724 *Biotechnol. Biofuels* **14**, 187 (2021).
725 61. Pick, A., Rühmann, B., Schmid, J. & Sieber, V. Novel CAD-like enzymes from Escherichia coli K-12 as
726 additional tools in chemical production. *Appl. Microbiol. Biotechnol.* **97**, 5815–5824 (2013).
727 62. Merchel Piovesan Pereira, B., Adil Salim, M., Rai, N. & Tagkopoulos, I. Tolerance to Glutaraldehyde in
728 Escherichia coli Mediated by Overexpression of the Aldehyde Reductase YqhD by YqhC. *Front. Microbiol.*
729 **12**, 680553 (2021).
730 63. Moreira De Gouveia, M. I., Daniel, J., Garrivier, A., Bernalier-Donadille, A. & Jubelin, G. Diversity of
731 ethanolamine utilization by human commensal Escherichia coli. *Res. Microbiol.* **174**, 103989 (2023).
732 64. Dellomonaco, C., Clomburg, J. M., Miller, E. N. & Gonzalez, R. Engineered reversal of the β-oxidation cycle
733 for the synthesis of fuels and chemicals. *Nature* **476**, 355–359 (2011).
734 65. Eram, M. & Ma, K. Decarboxylation of Pyruvate to Acetaldehyde for Ethanol Production by
735 Hyperthermophiles. *Biomolecules* **3**, 578–596 (2013).
736 66. Ho, K. K. & Weiner, H. Isolation and Characterization of an Aldehyde Dehydrogenase Encoded by the *aldB*
737 Gene of *Escherichia coli*. *J. Bacteriol.* **187**, 1067–1073 (2005).

738 67. Ravcheev, D. A., Moussu, L., Smajic, S. & Thiele, I. Comparative Genomic Analysis Reveals Novel
739 Microcompartment-Associated Metabolic Pathways in the Human Gut Microbiome. *Front. Genet.* **10**, 636
740 (2019).
741 68. Dank, A. *et al.* Bacterial Microcompartment-Dependent 1,2-Propanediol Utilization of Propionibacterium
742 freudenreichii. *Front. Microbiol.* **12**, 679827 (2021).
743 69. Chen, Y. *et al.* Associations of gut-flora-dependent metabolite trimethylamine-N-oxide, betaine and choline
744 with non-alcoholic fatty liver disease in adults. *Sci. Rep.* **6**, 19076 (2016).
745 70. Flores-Guerrero, J. L. *et al.* Circulating trimethylamine- *N* -oxide is associated with all-cause mortality in
746 subjects with nonalcoholic fatty liver disease. *Liver Int.* **41**, 2371–2382 (2021).
747 71. Wang, Z. *et al.* Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature* **472**,
748 57–63 (2011).
749 72. Roncal, C. *et al.* Trimethylamine-N-Oxide (TMAO) Predicts Cardiovascular Mortality in Peripheral Artery
750 Disease. *Sci. Rep.* **9**, 15580 (2019).
751 73. Martínez-del Campo, A. *et al.* Characterization and Detection of a Widely Distributed Gene Cluster That
752 Predicts Anaerobic Choline Utilization by Human Gut Bacteria. *mBio* **6**, e00042-15 (2015).
753 74. Goris, M. *et al.* Increased Thermostability of an Engineered Flavin-Containing Monooxygenase to
754 Remediate Trimethylamine in Fish Protein Hydrolysates. *Appl. Environ. Microbiol.* **89**, e00390-23 (2023).
755 75. Leimkühler, S. The biosynthesis of the molybdenum cofactors in *ESCHERICHIA COLI* . *Environ. Microbiol.* **22**,
756 2007–2026 (2020).
757 76. Kurth, J. M. *et al.* Methanogenic archaea use a bacteria-like methyltransferase system to demethoxylate
758 aromatic compounds. *ISME J.* **15**, 3549–3565 (2021).
759 77. Gendron, A. & Allen, K. D. Overview of Diverse Methyl/Alkyl-Coenzyme M Reductases and Considerations
760 for Their Potential Heterologous Expression. *Front. Microbiol.* **13**, 867342 (2022).

761 78. Thauer, R. K. Methyl (Alkyl)-Coenzyme M Reductases: Nickel F-430-Containing Enzymes Involved in
762 Anaerobic Methane Formation and in Anaerobic Oxidation of Methane or of Short Chain Alkanes.
763 *Biochemistry* **58**, 5198–5220 (2019).
764 79. Wagner, T., Koch, J., Ermler, U. & Shima, S. Methanogenic heterodisulfide reductase (HdrABC-MvhAGD)
765 uses two noncubane [4Fe-4S] clusters for reduction. *Science* **357**, 699–703 (2017).
766 80. Aziz, I. *et al.* Structural and mechanistic basis of the central energy-converting methyltransferase complex
767 of methanogenesis. *Proc. Natl. Acad. Sci.* **121**, e2315568121 (2024).
768 81. Almeida, A. *et al.* A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat.*
769 *Biotechnol.* **39**, 105–114 (2021).
770 82. Rosconi, F. *et al.* A bacterial pan-genome makes gene essentiality strain-dependent and evolvable. *Nat.*
771 *Microbiol.* **7**, 1580–1592 (2022).
772 83. Yu, M. K., Fogarty, E. C. & Eren, A. M. Diverse plasmid systems and their ecology across human gut
773 metagenomes revealed by PlasX and MobMess. *Nat. Microbiol.* **9**, 830–847 (2024).
774 84. O'Callaghan, J. L., Willner, D., Buttini, M., Huygens, F. & Pelzer, E. S. Limitations of 16S rRNA Gene
775 Sequencing to Characterize Lactobacillus Species in the Upper Genital Tract. *Front. Cell Dev. Biol.* **9**, 641921
776 (2021).
777 85. Rosero, J. A. *et al.* Reclassification of Eubacterium rectale (Hauduroy et al. 1937) Prévot 1938 in a new
778 genus Agathobacter gen. nov. as Agathobacter rectalis comb. nov., and description of Agathobacter
779 ruminis sp. nov., isolated from the rumen contents of sheep and cows. *Int. J. Syst. Evol. Microbiol.* **66**, 768–
780 773 (2016).
781 86. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the
782 tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).

783 87. Parks, D. H. *et al.* GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically
784 consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* **50**, D785–D794
785 (2022).
786 88. Medina, J. M., Fernández-López, R. & Crespo, J. Propionate Fermentative Genes of the Gut Microbiome
787 Decrease in Inflammatory Bowel Disease.
788 89. Louis, P., McCrae, S. I., Charrier, C. & Flint, H. J. Organization of butyrate synthetic genes in human colonic
789 bacteria: phylogenetic conservation and horizontal gene transfer: Organization of butyrate synthetic genes
790 in human colonic bacteria. *FEMS Microbiol. Lett.* **269**, 240–247 (2007).
791 90. Cui, Y. *et al.* Production of butyl butyrate from lignocellulosic biomass through Escherichia coli-Clostridium
792 beijerinckii G117 co-culture. *Process Biochem.* **128**, 58–67 (2023).
793 91. Gonzalez-Alba, J. M., Baquero, F., Cantón, R. & Galán, J. C. Stratified reconstruction of ancestral Escherichia
794 coli diversification. *BMC Genomics* **20**, 936 (2019).
795 92. Palomino, A. *et al.* Metabolic genes on conjugative plasmids are highly prevalent in *Escherichia coli* and can
796 protect against antibiotic treatment. *ISME J.* **17**, 151–162 (2023).
797 93. Abriouel, H. New insights into the role of plasmids from probiotic Lactobacillus pentosus MP-10 in Aloreña
798 table olive brine fermentation.
799 94. Pudlo, N. A. *et al.* Phenotypic and Genomic Diversification in Complex Carbohydrate-Degrading Human Gut
800 Bacteria. *mSystems* **7**, e00947-21 (2022).
801 95. Leinonen, R., Sugawara, H., Shumway, M., & on behalf of the International Nucleotide Sequence Database
802 Collaboration. The Sequence Read Archive. *Nucleic Acids Res.* **39**, D19–D21 (2011).
803 96. Poyet, M. *et al.* A library of human gut bacterial isolates paired with longitudinal multiomics data enables
804 mechanistic microbiome research. *Nat. Med.* **25**, 1442–1452 (2019).
805 97. Hoyles, L. *et al.* Molecular phenomics and metagenomics of hepatic steatosis in non-diabetic obese
806 women. *Nat. Med.* **24**, 1070–1080 (2018).

807 98. Mardinoglu, A. *et al.* An Integrated Understanding of the Rapid Metabolic Benefits of a Carbohydrate-
808 Restricted Diet on Hepatic Steatosis in Humans. *Cell Metab.* **27**, 559-571.e5 (2018).
809 99. Qin, N. *et al.* Alterations of the human gut microbiome in liver cirrhosis. *Nature* **513**, 59–64 (2014).
810 100. BBMap.
811 101. FastQC.
812 102. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools
813 and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
814 103. Togo, A. H. *et al.* Description of Mediterraneibacter massiliensis, gen. nov., sp. nov., a new genus
815 isolated from the gut microbiota of an obese patient and reclassification of Ruminococcus faecis,
816 Ruminococcus lactaris, Ruminococcus torques, Ruminococcus gnavus and Clostridium glycyrrhizinilyticum
817 as Mediterraneibacter faecis comb. nov., Mediterraneibacter lactaris comb. nov., Mediterraneibacter
818 torques comb. nov., Mediterraneibacter gnavus comb. nov. and Mediterraneibacter glycyrrhizinilyticus
819 comb. nov. *Antonie Van Leeuwenhoek* **111**, 2107–2128 (2018).
820 104. Murotomi, K., Tourlousse, D. M., Hamajima, M. & Sekiguchi, Y. Complete genome sequence of
821 *Roseburia faecis* M72/1^{T}. *Microbiol. Resour. Announc.* **14**, e01253-24 (2025).
822 105. Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic*
823 *Acids Res.* **48**, D445–D453 (2020).
824 106. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in
825 Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
826 107. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the
827 Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
828 108. HMMER.
829 109. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes.

830 110. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND.
831 *Nat. Methods* **18**, 366–368 (2021).
832 111. Treiber, M. L., Taft, D. H., Korf, I., Mills, D. A. & Lemay, D. G. Pre- and post-sequencing
833 recommendations for functional annotation of human fecal metagenomes. *BMC Bioinformatics* **21**, 74
834 (2020).
835 112. Haft, D. H. *et al.* RefSeq and the prokaryotic genome annotation pipeline in the age of metagenomes.
836 *Nucleic Acids Res.* **52**, D762–D769 (2024).
837 113. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC*
838 *Bioinformatics* **11**, 119 (2010).
839 114. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
840 115. ggplot2.
841 116. ggpubr.
842 117. pheatmap.
843 118. rstatix.
844