

1 **Novel genetic insights into the roles of amino acids in metabolic dysfunction-associated steatotic liver**
2 **disease**

3 Jun Liu¹, Yuxuan Chen², Jin Qian³, Ri Cui⁴, Ayse Demirkan⁵, Yihu Zheng³

4

5 **Affiliations**

6 1. Nuffield Department of Population Health, University of Oxford, UK

7 2. The First School of Medicine, Wenzhou Medical University, China

8 3. Department of General Surgery, the First Affiliated Hospital of Wenzhou Medical University, China

9 4. School of Pharmaceutical Sciences, Wenzhou Medical University, China

10 5. People-centred AI institute & Section of Statistical Multi-Omics, Department of Clinical & Experimental

11 Medicine, School of Biosciences & Medicine, University of Surrey, Guildford, UK

12

13 **Corresponding authors**

14 Jun Liu: jun.liu@ndph.ox.ac.uk ;

15 Yihu Zheng: zhengyihu@wmu.edu.cn

16

17 **Keywords:** Metabolic dysfunction-associated steatotic liver disease; amino acids; valine; leucine; alanine;

18 branched-chain amino acids; phenylalanine; Mendelian randomization

19

20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

Abstract

Background

Previous research has suggested potential links between amino acids and metabolic dysfunction-associated steatotic liver disease (MASLD), but the precise roles of amino acids in MASLD development are not well understood. This study aimed to obtain insights into the relationships between circulating amino acids and MASLD.

Methods

Utilizing data from the UK Biobank, we examined the observational associations of ten amino acids with MASLD in a cohort of 72,626 MASLD cases and 128,102 controls. Bi-directional two-sample Mendelian randomization (MR) was conducted using genome-wide association study data to investigate the causal relationships between amino acids and MASLD. Multiple MR methods comprising MR-Egger and MR-PRESSO were applied to assess pleiotropy and heterogeneity, and multivariable MR was conducted to evaluate the impacts of body mass index (BMI) on these associations. Survival analysis assessed the link between baseline amino acid levels and the risk of major outcomes.

Results

We identified nine amino acids significantly associated with MASLD in the observational study. The genetic predisposition towards higher leucine (odds ratio (OR) [95% confidence interval (CI)]: 2.1 [1.4, 3.2]), valine (OR [95% CI]: 1.8 [1.3, 2.7]), and alanine (OR [95% CI]: 1.4 [1.1, 1.8]) levels were significantly associated with MASLD. By contrast, the genetic predisposition for increased MASLD risk was significantly associated with phenylalanine ($\beta = 0.05$, $p = 4.0 \times 10^{-4}$). Further analysis showed that valine may mediate the association between BMI and MASLD, and may also have an exclusive effect on MASLD in addition to the effect of obesity ($\beta = 1.3$, $p = 1.9 \times 10^{-4}$). Elevated phenylalanine levels in MASLD patients were linked with an increased risk of metabolic dysfunction-associated steatohepatitis (MASH), hepatocellular carcinoma, cirrhosis, heart failure, stroke, and mortality.

48 **Conclusion**

49 We found genetic associations between circulating branched-chain amino acids, particularly leucine and
50 valine, and MASLD, independent of obesity. Phenylalanine was identified as a potential biomarker for
51 MASLD prognostic complications. These results highlight the importance of amino acid metabolism in
52 MASLD as well as suggest new possibilities for research and therapeutic intervention.

53 Introduction

54 Metabolic dysfunction-associated steatotic liver disease (MASLD), previously known as non-alcoholic
55 fatty liver disease, is a multifactorial condition with a global prevalence of approximately 25% [1]. The
56 progressive form of MASLD, metabolic dysfunction-associated steatohepatitis (MASH), can lead to severe
57 liver complications, and thus MASLD is a leading cause of liver-related mortality. The disease is closely
58 linked to obesity and metabolic syndrome, where it manifests as hepatic insulin resistance and is
59 associated with comorbidities such as type 2 diabetes and cardiovascular diseases [1-2]. Amino acids,
60 particularly branched-chain amino acids (BCAAs), are related to obesity and essential in metabolic
61 regulation. They have been implicated in the pathogenesis of various metabolic disorders. The BCAAs
62 consist of valine, leucine, and isoleucine are essential amino acids obtained from dietary sources.
63 Previous studies have suggested roles for BCAAs in the onset of type 2 diabetes and cardiovascular
64 diseases, and emerging evidence suggests their involvement in MASLD [3-5]. Increased levels of fasting
65 plasma BCAAs are partly mediate the association between MASLD and incident type 2 diabetes [6].

66
67 Despite the fact that observational studies have frequently reported higher plasma concentrations of
68 amino acids, particularly BCAAs, in cases of MASLD [7], the nature of their relationship is unclear.
69 Mendelian randomization (MR) uses genetic variations as instrumental variables to investigate potential
70 relationships and their directions, thereby providing an alternative approach that can help reduce the
71 effects of confounding factors and reverse causation. A study based on MR linked MASLD with increased
72 plasma tyrosine levels, suggesting amino acids other than BCAA may also be relevant [8]. Furthermore,
73 MR findings indicate associations between higher levels of alanine and glutamine and the risk of MASLD
74 [9], thereby contributing to our understanding of the metabolic underpinnings of MASLD. However, MR
75 methodology is improved by the day, so as the strength of instrumental variables detected by larger
76 association studies. Additionally, there is a gap in literature as most of the MR research were not
77 supported by individual data or follow-up analysis in addition to the MR results, such as the role of body
78 mass index (BMI) in the association and the effects of biomarkers on severe outcomes related to MASLD.
79 Therefore, further research is needed to clarify these relationships.

80

81 In the present study, we aimed to elucidate the genetic underpinnings of the relationships between
82 circulating amino acids and MASLD by employing MR. By utilizing the extensive data in the UK Biobank
83 and advanced MR techniques, we assessed the associations between specific amino acids and the risk
84 of MASLD and major outcomes among patients, as well as the potential mediating role of BMI in these
85 relationships, as BMI is the most important risk factor for MASLD and it is also strongly associated with
86 amino acids. Our findings provide a more definitive understanding of the roles of amino acids in MASLD
87 to potentially inform future therapeutic strategies and contribute to the prevention of this increasingly
88 prevalent disease.

89

90 **Methods**

91 *Study population*

92 Our study was conducted based on the UK Biobank, which is a large-scale cohort study that includes over
93 500,000 participants who were aged between 37 and 73 years during the recruitment period (2006 to
94 2010). These participants were registered with the UK National Health Service and from 22 assessment
95 centers across England, Wales, and Scotland. Standardized procedures were used for data collection,
96 including a wide range of questionnaires, whole-body magnetic resonance imaging, physical and
97 anthropometric measurements, clinical biomarkers, metabolite measurements, genotype data, and
98 electronic health record linkage. All participants provided electronically signed informed consent, and
99 the study was approved by the North West Multi-centre Research Ethics Committee, Patient Information
100 Advisory Group, and Community Health Index Advisory Group. The current study is part of UK Biobank
101 project 61054. Further details about the UK Biobank are available online.

102

103 We excluded participants with non-European ancestry, excessive alcohol consumption (regularly
104 drinking > 60 g/day for males or > 50 g/day for females), patients with other underlying conditions [10],
105 and samples which were not measured by the nuclear magnetic resonance (NMR) spectroscopy
106 metabolic panel at baseline. After applying these exclusion criteria, we retained 200,728 participants for
107 the current study.

108

109 *Definitions of MASLD, major outcomes, and covariates*

110 We identified cases of MASLD based on fatty liver index ≥ 60 and presence of at least one of five
111 cardiometabolic criteria reported previously by Rinella ME, et al, at baseline [11]. Other underlying
112 conditions of liver diseases and major outcomes (stroke, heart failure, myocardial infarction, MASH,
113 cirrhosis and hepatocellular carcinoma, and mortality) were defined based on self-reported, hospital
114 inpatient records, or death registration up to November 2022. We considered the earliest documented
115 instance of a condition as the official diagnosis date. Prevalent cases referred to participants who had a
116 recorded diagnosis date on or before their initial assessment, or those who self-reported a condition at
117 that time, and were excluded from the survival analysis. The censor date was set as the earliest date of
118 the first recorded event, the date of death, or the latest recorded date. Alcohol intake was quantified
119 based on a questionnaire about average weekly alcohol consumption. Medication usage was obtained
120 from a touchscreen questionnaire at baseline. The definitions of other covariates, including BMI, blood
121 biochemical biomarkers, fasting time, type 2 diabetes, hypertension, medication information, smoking
122 status, smoking pack-years, education, and physical activity, were presented in a previous study [10].

123
124 We utilized the R package *missRanger*, which employs chained random forests, for rapid imputation of
125 missing data necessary to define FLI and MASLD. The imputation process incorporated a range of
126 baseline information, including age, sex, smoking status, pack-years of smoking, frequency of alcohol
127 consumption, physical activity levels as categorized by the International Physical Activity Questionnaire
128 (IPAQ), ethnicity, BMI, education level, blood pressure, and waist-hip ratio. To summarize, the extensive
129 data matrix underwent imputation with up to ten iterations of chained equations and 200 trees, with
130 the process being weighted according to the count of existing non-missing values. During the predictive
131 mean matching phase, three potential non-missing values were chosen as candidates for each missing
132 entry.

133
134 *Measurement of amino acids*

135 Baseline amino acids were measured in plasma using the high-throughput $^1\text{H-NMR}$ metabolomics
136 platform (Nightingale Health, Helsinki, Finland) in a randomly selected subset of 274,124 samples from

137 the UK Biobank [12]. This platform uses a standardized protocol for sample quality control, sample
138 preparation, data storage, and automated spectral analyses, which was previously described in detail
139 [13]. To prepare the data for analysis, we performed natural logarithm transformation of the ten available
140 amino acid variables (i.e., valine, leucine, isoleucine, histidine, glutamine, glycine, tyrosine, alanine,
141 phenylalanine, and total BCAA,) reported by the platform. Metabolite values were rank-based inverse
142 normal transformed before running any analysis.

143

144 *Statistical analysis*

145 All analyses were performed using R statistical software (version 4.0.3) and two-tailed tests were
146 conducted.

147

148 To investigate the associations between amino acids and MASLD, we performed logistic regression
149 analysis for each separate amino acid. The Cox proportional hazards model was used to assess the
150 associations between baseline amino acids levels and the risk of incident major outcomes during follow-
151 up in MASLD patients. We used three models to adjust for potential confounding factors: model 1
152 included age, sex, and fasting time; model 2 included the covariates in model 1 as well as common
153 lifestyle factors, including smoking status, number of pack-years of smoking, grams of alcohol
154 consumption per week, education, and physical activities; and model 3 included the covariates in model
155 2 and BMI. Multiple testing was considered with adjustment of the independent equivalent number of
156 tests by Matrix Spectral Decomposition (MSD) method [14] .

157

158 For the amino acids that had significant observational associations with MASLD, we further explored
159 their associations by conducting two-sample bi-directional MR using the *TwoSampleMR* R package. The
160 MR analyses were predicated on three assumptions: (1) the genetic variants were associated with
161 exposure; (2) the genetic instruments were not associated with the outcomes through confounding
162 factors; and (3) the genetic instruments did not have direct effects on the outcome, but potentially only
163 through exposure (Figure 1).

164

165 We used data from the largest available genome-wide association study (GWAS) for amino acids from
166 the UK Biobank, which included 115,082 European participants [15]. The GWAS of MASLD was obtained
167 from the study by Ghodsian et al., which included 8,434 MASLD patients and 770,180 controls [16].
168 Instrumental variables were selected using the independent genetic determinants ($p < 5 \times 10^{-8}$) of
169 exposure (either amino acid or MASLD) with a clumping process to exclude single nucleotide
170 polymorphisms (SNPs) in linkage disequilibrium (threshold: $R^2 > 0.001$ and distance $< 10,000$ kb) [17].
171 The inverse variance weighted (IVW) method under random effects was used as the primary statistical
172 method. We also calculated the F statistics for weak instrument bias using the formula: $F = [(N - K -$
173 $1)/K] * [R^2 / (1 - R^2)]$, where N is the sample size, K is the number of IVs, and R^2 is the proportion of the
174 variability of the exposure explained by IVs.

175
176 Sensitivity analysis was performed for the significant MR results with further excluding potentially
177 ambiguous SNPs due to reference strand or palindromic SNPs. Cochran's Q-test was conducted to
178 assess the heterogeneity across the individual effect estimates derived from each genetic variant. We
179 further tested our hypothesis using other MR methods; weighted median regression [18], MR-Egger
180 regression [19], weighted mode [20], simple mode [20], and MR-Pleiotropy RESidual Sum and Outlier
181 (MR-PRESSO) [21].

182
183 For the amino acids potentially on the pathway to MASLD, we further performed multivariable MR to
184 explore the role of BMI in the association of these amino acids and MASLD, using the R package
185 TwoSampleMR with BMI GWAS from Yengo et al., which included 681,275 European participants [22].
186 The same GWAS sources and pipeline were applied for the MR of BMI and amino acids.

187
188 **Results**
189 In the current study, we included 72,626 pure MASLD cases and 128,102 controls. As shown in Table 1,
190 patients with MASLD exhibited increased level metabolic risk factors, including older age, a higher
191 proportion of males, higher BMI, larger waist circumference, higher glucose levels, higher HbA1c levels,
192 higher triglycerides, and higher liver enzymes. Remarkably, these patients had less favorable lifestyles,

193 such as higher rates of smoking and alcohol consumption, lower levels of education, and lower physical
194 activity.

195

196 *Observational associations between plasma amino acids and MASLD*

197 We compared the levels of ten plasma amino acids between MASLD cases and controls across three
198 models (Table 2). Nine out of the ten amino acids were statistically significantly associated with MASLD
199 across all three models ($p < 6.25 \times 10^{-3}$), including six amino acids (valine, leucine, isoleucine, tyrosine,
200 alanine, and phenylalanine) and total BCAA that had positive associations with MASLD, and glutamine
201 and glycine, which had negative associations with MASLD. The effect estimates for these nine amino
202 acids with MASLD were very similar under model 1 and model 2, thereby suggesting that common
203 lifestyle factors had limited effects on their associations. The absolute effect estimates for these nine
204 amino acid variables on MASLD generally decreased after adjustment for BMI on top of model 2 (model
205 3). The association between histidine and MASLD was significant in three models, but the direction of
206 effect estimates shifted from negative association (beta = -0.34 in model 1 and beta = -0.03 in model 2)
207 to positive association (beta = 0.03) after adjusting for BMI, which suggests that their negative
208 associations may be masked by the risk effect of BMI on MASLD.

209

210 *Bi-directional MR of amino acids and MASLD*

211 For the nine amino acids that had robust significant associations with MASLD, we performed two-sample
212 bi-directional MR based on the previously published GWAS of MASLD and the GWAS of the UK Biobank
213 [15-16] (Figure 2). The number of genetic loci selected in the instrumental variables for amino acids
214 ranged from seven for isoleucine and phenylalanine to as many as 35 for glycine, with F statistics ranging
215 from 62.2 for leucine to 603.9 for glycine. We found that the genetic predisposition towards increased
216 levels of leucine (odds ratio (OR) [95% confidence interval (CI)]: 2.1 [1.4, 3.2]), valine (OR [95% CI]: 1.8
217 [1.3, 2.7]), and alanine (OR [95% CI]: 1.4 [1.1, 1.8]) were statistically significantly associated with
218 increased risk of MASLD ($p < 3.6 \times 10^{-3}$).

219

220 Four genetic loci were selected as the instrumental variables for MASLD, i.e., rs28601761, rs3747207,
221 rs429358, and rs73001065, which had F statistics of 13.4. The F statistics for MASLD were not as strong
222 as those for amino acids, but we still detected statistically significant differences between the genetic
223 predisposition toward MASLD and increased levels of tyrosine ($\beta = 0.15$, $p = 2.0 \times 10^{-3}$) as well as
224 phenylalanine ($\beta = 0.047$, $p = 4.0 \times 10^{-4}$) ($p < 3.6 \times 10^{-3}$).

225

226 *Sensitivity analysis based on significant MR findings*

227 For the significant MR analysis results, including three amino acids (valine, leucine, and alanine) relative
228 to MASLD and MASLD relative to two amino acids (phenylalanine and tyrosine), we explored the
229 potential heterogeneity and pleiotropic effect in MR (Figure 3) using additional MR methods. Significant
230 association of between genetic predisposition to elevated valine, leucine, and alanine were confirmed
231 by the weighted mean method, while for the other methods significance varied with different
232 assumptions while directions of the effect estimates were stable (Figure 3A). On the other hand, the
233 effect of MASLD on amino acids phenylalanine and tyrosine was significant for all methods tested, except
234 MR-Egger (Figure 3B). Sensitivity analysis excluding potentially ambiguous and/or palindromic SNPs in
235 the instrumental variables did not affect the significance of the MR analysis results, except the
236 association between MASLD and tyrosine shifted to nominal significance ($p = 0.022$) when we excluded
237 rs28601761 (Figure 3C-D).

238

239 Furthermore, Cochran's Q-test demonstrated significant heterogeneity among the SNPs included in the
240 instrumental variables for alanine with MASLD ($p = 5.0 \times 10^{-3}$) and in the instrumental variables for MASLD
241 with tyrosine ($p = 8.36 \times 10^{-9}$; $p < 0.01$). MR-PRESSO was also applied as another method to detect
242 heterogeneity among the SNPs used in the instrumental variables and it consistently detected significant
243 heterogeneity through the global test for alanine with MASLD ($p = 4.0 \times 10^{-3}$, $p < 0.01$). The distortion test
244 detected no difference in the MR results before and after excluding outliers for the link between alanine
245 and MASLD ($p > 0.01$). Moreover, no significant pleiotropic effect was detected by MR-Egger regression
246 ($p > 0.01$). The MR results for leucine, valine, and phenylalanine were robust according to heterogeneity
247 and pleiotropy tests.

248

249 *Role of BMI in the associations between amino acids and MASLD*

250 BMI is the most important risk factor for MASLD and it is also strongly associated with amino acids, so it
251 was essential to explore its role in the associations between amino acids and MASLD. First, we
252 investigated the potential role of BMI in the associations of the three amino acids (leucine, valine, and
253 alanine) and MASLD by MR (Table 3). We found that the genetic predisposition towards higher BMI was
254 strongly associated with both BCAAs, i.e., leucine and valine (beta = 0.19, $p = 1.84 \times 10^{-27}$ for leucine; beta
255 = 0.23, $p = 1.36 \times 10^{-37}$ for valine), through IVW. MR-PRESSO showed that the heterogeneity among the
256 instrumental variables did not affect their significance ($p > 0.05$) and no pleiotropic effect was detected.
257 The association of genetic predisposition towards higher BMI and alanine was borderline significant ($p =$
258 0.02, $p < 0.017$).

259

260 After considering the effect of BMI on the associations of leucine, valine, and alanine with MASLD by
261 multivariable MR (Table 4), a significant association remained for valine (beta = 1.28, $p = 1.9 \times 10^{-4}$), which
262 suggested potential pathways from valine to MASLD independent of the obesity effect. A high beta value
263 was also obtained for leucine (beta = 2.9). No significance was detected for alanine and MASLD after
264 adjusting for the effect of BMI ($p = 0.31$), thereby suggesting that the association between alanine and
265 MASLD was likely to be fully mediated by the effect of obesity.

266

267 *Roles of phenylalanine and tyrosine in major outcomes of MASLD patients*

268 Phenylalanine and tyrosine were found to be associated with the genetic predisposition toward MASLD,
269 and thus the alterations in them may result to the progression of MASLD. Therefore, we further explored
270 their associations with the incidence of major outcomes in MASLD patients (Figure 4). The results
271 showed that baseline phenylalanine was significantly associated with increased risk of MASH (hazard
272 ratio (HR) [95% CI]: 1.3 [1.16–1.45]), hepatocellular carcinoma (HR [95% CI]: 1.56 [1.32–1.85]), cirrhosis
273 (HR [95% CI]: 1.14 [1.08–1.2]), heart failure (HR [95% CI]: 1.13 [1.09–1.14]), stroke (HR [95% CI]: 1.09
274 [1.04–1.14]), and mortality (HR [95% CI]: 1.06 [1.04–1.09]), and not affected by lifestyle factors or BMI
275 ($p < 7.1 \times 10^{-3}$). The associations of tyrosine and increased risk of liver related outcomes with MASH (HR

276 [95% CI]: 1.71 [1.52–1.92]), hepatocellular carcinoma (HR [95% CI]: 2.04 [1.73–2.42]), and cirrhosis (HR
277 [95% CI]: 1.2 [1.14–1.27]) were consistent across the adjusted models, but not associated with any
278 cardiovascular events or mortality after accounting for BMI.

279

280 **Discussion**

281 In the present study, for the first time, we established a robust genetic link between valine, leucine, and
282 phenylalanine with MASLD using a MR approach. We found that valine and leucine may mediate the
283 relationship between BMI and MASLD but it could also be involved in unique pathways leading to MASLD
284 in addition to the effect of obesity, thereby suggesting that BCAAs could potentially serve as targets for
285 MASLD prevention or treatment. In addition, we identified phenylalanine as a potential novel outcome
286 of MASLD, which could be developed as an early detection biomarker for severe liver and cardiovascular
287 outcomes in MASLD patients.

288

289 Following our observational findings which indicated a strong correlation between various amino acids
290 and MASLD, we provided genetic evidence of associations between genetic predisposition of leucine,
291 valine, and alanine with MASLD, and between genetic predisposition of MASLD with tyrosine and
292 phenylalanine. Our results validated previous findings by Zhao et al. [9], who reported that higher plasma
293 alanine levels alter a higher risk of MASLD by MR approach. In addition to that, we obtained further
294 evidence that their association is likely to be mediated by obesity as adjustment for BMI by multivariable
295 MR resulted in a non-significant association ($p = 0.31$). Moreover, our demonstration of an association
296 between MASLD with tyrosine is consistent with previous findings by Gobeil et al. [8] who used different
297 data sources for the MR analysis. Moreover, our results also provide extra information about the effect
298 of tyrosine on the risk of incident liver outcomes (MASH, hepatocellular carcinoma, and cirrhosis) in
299 MASLD patients, indicating that as a result of MASLD development, alterations in tyrosine may play an
300 important role in the liver complications of MASLD. Further MR analyses of tyrosine and liver outcomes
301 in MASLD patients, which are not yet available, could provide more evidence for the observational
302 prospective cohort.

303

304 Notably, the associations between leucine and valine with MASLD and the association of MASLD with
305 phenylalanine obtained by the MR approach were determined for the first time with robust results
306 through various sensitivity analyses. Moreover, we also detected a complex interaction between BMI,
307 valine, and MASLD. BMI led to the incidence of MASLD with a significant mediating effect from valine
308 and leucine, and valine also had an exclusive effect on MASLD other than through obesity. These findings
309 suggest that BCAAs, especially valine, could have important roles in the further development of
310 prevention and treatment strategies for MASLD. Currently, the mechanism that allows plasma BCAAs to
311 cause MASLD is not well understood. However, it has been shown that overconsumption of valine
312 triggers MASLD in laying hens by stimulating fatty acid synthesis in the liver [23]. Experimental studies
313 have shown that elevated BCAAs can damage the oxidation of carbon substrates through the
314 tricarboxylic acid cycle (TCA cycle) to result in mitochondrial dysfunction and participation in MASLD [24].
315 In addition, leucine promotes MASLD by modulating AMP-activated protein kinase through myostatin to
316 lead to hepatocyte triglyceride accumulation [25]. Moreover, aminoacylation, as a novel amino acid
317 modification, allows BCAAs to exert their regulatory roles in cellular functions through the effect of lysine
318 aminoacylation on specific substrate proteins catalyzed by aminoacyl-tRNA synthetase [26]. In addition,
319 many factors can affect the concentrations of plasma BCAAs, including exogenous BCAA uptake, adipose
320 tissue, and skeletal muscle, which may be a part of personalized therapeutic routes for MASLD
321 management involving BCAAs.

322
323 In addition to the important roles of BCAAs in the pathway to MASLD, it should be noted that we
324 identified phenylalanine as a new biomarker for MASLD. Previously, the correlation between non-obese
325 individuals with MASLD and phenylalanine levels was discovered using metabolomics approaches [27].
326 However, no clinical data with a large sample size are available to support this correlation and causal
327 inferences were not found by MR analysis. In the present study, for the first time, we found that
328 alterations in blood phenylalanine levels were associated with the genetic predisposition toward MASLD.
329 Moreover, we provided evidence based on prospective cohort data that the baseline phenylalanine level
330 in MASLD patients may be an early biomarker for liver and cardiovascular major outcomes and mortality.
331 Our results are consistent with those obtained in previous clinical studies. The plasma phenylalanine

332 increased according to the severity of the disease from steatosis to non-alcoholic steatohepatitis
333 compared with normal control [28-29]. Moreover, in a prospective cohort study, the plasma
334 phenylalanine concentration correlated with hepatocellular carcinoma occurrence in liver cirrhotic
335 patients after following for the next 3 years [30].

336
337 The present study also had some limitations. First, the GWAS summary statistics for MASLD and amino
338 acids involved overlapping participants, but the overlap for MASLD cases was only 19.7% (1,644 in 8,434),
339 which was unlikely to have affected our major findings [31]. Second, in case of SNP missingness, we did
340 not opt for replacing them with proxy SNPs for the instrumental variables of exposures, which may have
341 reduced the explained variance for the instrumental variables on exposure. However, this only increased
342 the likelihood of false-negative findings and would not have affected our significant results. Third, we
343 only included ten amino acids in the study. However, individual data were available for these amino acids
344 to support their observational associations with MASLD and MASLD complications. Finally, more tissue-
345 based studies are required to understand the mechanisms that underlie the relationships between
346 plasma BCAAs and MASLD.

347
348 In conclusion, we established a link between circulating BCAAs and MASLD using a MR approach, thereby
349 suggesting that BCAAs could potentially serve as targets for the prevention or treatment of MASLD. We
350 also highlighted the complex interactions between BMI, BCAAs, and MASLD, and the important roles of
351 BCAAs in the development of prevention and treatment strategies for MASLD. Furthermore, we
352 identified alteration in phenylalanine as a novel potential outcome of MASLD, which could be developed
353 as a biomarker for the early detection of MASLD and predicting the incidence of major outcomes in
354 MASLD patients. Our study significantly contributes to understanding the interplay between amino acids
355 and MASLD to provide a foundation for future research and the potential development of therapeutic
356 strategies. Further tissue-based studies are required to elucidate the mechanisms that underlie their
357 relationships.

358

359 **Conflict of interest statement:** The authors declare no competing interests.

360 Author contributions: J.L. and Y.Z. conceived and designed the current study. J.L., Y.C., and J.Q. performed
361 analyses. J.L, Y.C., and Y.Z. prepared the manuscript. R.C, A.D. and Y.Z. reviewed and edited the
362 manuscript. All authors read and approved the manuscript.

363
364 **Data availability:** UK Biobank data are publicly available to bona fide researchers upon application at
365 <http://www.ukbiobank.ac.uk/using-the-resource/>. Publicly available summary statistics are obtained
366 from <https://gwas.mrcieu.ac.uk/> and <https://www.ebi.ac.uk/gwas/>. Other sources of data or web
367 sources were clarified in the methods.

368
369 **Acknowledgement:** This research was conducted using data from UK Biobank, a major biomedical
370 database (<https://www.ukbiobank.ac.uk/>) via application no. 61054. We thank the participants,
371 contributors, clinicians, and researchers for making data available for this study. J.L. is supported by a
372 Novo Nordisk Postdoctoral Fellowship Programme run in partnership with the University of Oxford. Y. Z
373 is supported by Discipline Cluster of Oncology, Wenzhou Medical University, China (No.z2-2023024).

374
375

376 **References**

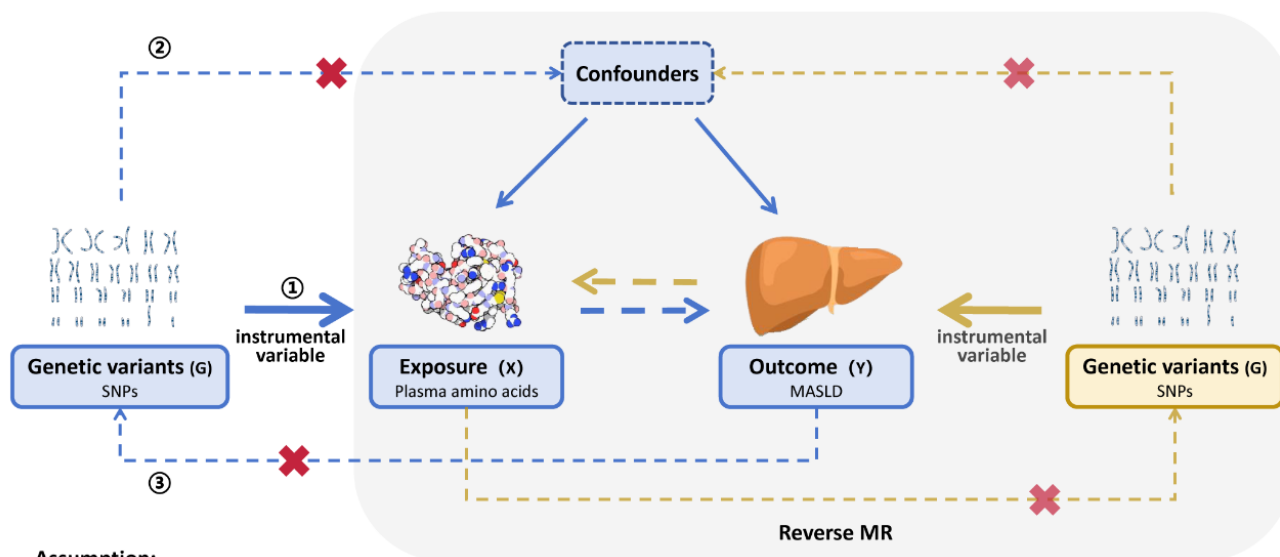
- 377
- 378 [1] Loomba R, Friedman SL, Shulman GI. Mechanisms and disease consequences of nonalcoholic fatty
379 liver disease. *Cell*. 2021 May 13;184(10):2537-2564.
 - 380 [2] Byrne CD, Targher G. NAFLD as a driver of chronic kidney disease. *J Hepatol*. 2020 Apr;72(4):785-801.
 - 381 [3] Ramzan I, Ardavani A, Vanweert F, Mellett A, Atherton PJ, Idris I. The Association between Circulating
382 Branched Chain Amino Acids and the Temporal Risk of Developing Type 2 Diabetes Mellitus: A Systematic
383 Review & Meta-Analysis. *Nutrients*. 2022 Oct 20;14(20):4411.
 - 384 [4] Doestzada M, Zhernakova DV, C L van den Munckhof I, Wang D, Kurilshikov A, Chen L, Bloks VW, van
385 Faassen M, Rutten JHW, Joosten LAB, Netea MG, Wijmenga C, Riksen NP, Zhernakova A, Kuipers F, Fu J.
386 Systematic analysis of relationships between plasma branched-chain amino acid concentrations and

- 387 cardiometabolic parameters: an association and Mendelian randomization study. *BMC Med.* 2022 Dec
388 15;20(1):485.
- 389 [5] Guo F, Chen R, Kong L, Wei P, Liu Z, Wang X, Hao H, Lu Y, Hu W. Effects of serum branched-chain amino
390 acids on nonalcoholic fatty liver disease and subsequent cardiovascular disease. *Hepatol Int.* 2022
391 Dec;16(6):1424-1434.
- 392 [6] van den Berg EH, Flores-Guerrero JL, Gruppen EG, de Borst MH, Wolak-Dinsmore J, Connelly MA,
393 Bakker SJL, Dullaart RPF. Non-Alcoholic Fatty Liver Disease and Risk of Incident Type 2 Diabetes: Role of
394 Circulating Branched-Chain Amino Acids. *Nutrients.* 2019 Mar 26;11(3):705.
- 395 [7] Grzych G, Vonghia L, Bout MA, Weyler J, Verrijken A, Dirinck E, Chevalier Curt MJ, Van Gaal L, Paumelle
396 R, Francque S, Tailleux A, Haas JT, Staels B. Plasma BCAA Changes in Patients With NAFLD Are Sex
397 Dependent. *J Clin Endocrinol Metab.* 2020 Jul 1;105(7):dgaa175.
- 398 [8] Gobeil É, Maltais-Payette I, Taba N, Brière F, Ghodsian N, Abner E, Bourgault J, Gagnon E, Manikpurage
399 HD, Couture C, Mitchell PL, Mathieu P, Julien F, Corbeil J, Vohl MC, Thériault S, Esko T, Tchernof A,
400 Arsenault BJ. Mendelian Randomization Analysis Identifies Blood Tyrosine Levels as a Biomarker of Non-
401 Alcoholic Fatty Liver Disease. *Metabolites.* 2022 May 13;12(5):440.
- 402 [9] Zhao J, Zeng J, Zhu C, Li X, Liu D, Zhang J, Li F, Targher G, Fan JG. Genetically predicted plasma levels
403 of amino acids and metabolic dysfunction-associated fatty liver disease risk: a Mendelian randomization
404 study. *BMC Med.* 2023 Nov 28;21(1):469.
- 405 [10] Liu J, Hu S, Chen L, Daly C, Prada Medina CA, Richardson TG, Traylor M, Dempster NJ, Mbasu R,
406 Monfeuga T, Vujkovic M, Tsao PS, Lynch JA, Voight BF, Chang KM, Million VA, Cobbold JF, Tomlinson JW,
407 van Duijn CM, Howson JMM. Profiling the genome and proteome of metabolic dysfunction-associated
408 steatotic liver disease identifies potential therapeutic targets. *medRxiv [Preprint].* 2023 Nov
409 30:2023.11.30.23299247.
- 410 [11] Rinella ME, Lazarus JV, Ratziu V, Francque SM, Sanyal AJ, Kanwal F, Romero D, Abdelmalek MF, Anstee
411 QM, Arab JP, Arrese M, Bataller R, Beuers U, Boursier J, Bugianesi E, Byrne CD, Castro Narro GE,
412 Chowdhury A, Cortez-Pinto H, Cryer DR, Cusi K, El-Kassas M, Klein S, Eskridge W, Fan J, Gawrieh S, Guy
413 CD, Harrison SA, Kim SU, Koot BG, Korenjak M, Kowdley KV, Lacaille F, Loomba R, Mitchell-Thain R,
414 Morgan TR, Powell EE, Roden M, Romero-Gómez M, Silva M, Singh SP, Sookoian SC, Spearman CW,

- 415 Tiniakos D, Valenti L, Vos MB, Wong VW, Xanthakos S, Yilmaz Y, Younossi Z, Hobbs A, Villota-Rivas M,
416 Newsome PN; NAFLD Nomenclature consensus group. A multisociety Delphi consensus statement on
417 new fatty liver disease nomenclature. *J Hepatol.* 2023 Dec;79(6):1542-1556.
- 418 [12] Julkunen H, Cichońska A, Tiainen M, Koskela H, Nybo K, Mäkelä V, Nokso-Koivisto J, Kristiansson K,
419 Perola M, Salomaa V, Jousilahti P, Lundqvist A, Kangas AJ, Soininen P, Barrett JC, Würtz P. Atlas of plasma
420 NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank. *Nat Commun.* 2023
421 Feb 3;14(1):604.
- 422 [13] Würtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M. Quantitative Serum
423 Nuclear Magnetic Resonance Metabolomics in Large-Scale Epidemiology: A Primer on -Omic
424 Technologies. *Am J Epidemiol.* 2017 Nov 1;186(9):1084-1096.
- 425 [14] Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix.
426 *Heredity (Edinb).* 2005 Sep;95(3):221-227.
- 427 [15] Richardson TG, Leyden GM, Wang Q, Bell JA, Elsworth B, Davey Smith G, Holmes MV. Characterising
428 metabolomic signatures of lipid-modifying therapies through drug target mendelian randomisation.
429 *PLoS Biol.* 2022 Feb 25;20(2):e3001547.
- 430 [16] Ghodsian N, Abner E, Emdin CA, Gobeil É, Taba N, Haas ME, Perrot N, Manikpurage HD, Gagnon É,
431 Bourgault J, St-Amand A, Couture C, Mitchell PL, Bossé Y, Mathieu P, Vohl MC, Tchernof A, Thériault S,
432 Khera AV, Esko T, Arsenault BJ. Electronic health record-based genome-wide meta-analysis provides
433 insights on the genetic architecture of non-alcoholic fatty liver disease. *Cell Rep Med.* 2021 Nov
434 3;2(11):100437.
- 435 [17] Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, Laurin C, Burgess S, Bowden J,
436 Langdon R, Tan VY, Yarmolinsky J, Shihab HA, Timpson NJ, Evans DM, Relton C, Martin RM, Davey Smith
437 G, Gaunt TR, Haycock PC. The MR-Base platform supports systematic causal inference across the human
438 phenome. *Elife.* 2018 May 30;7:e34408.
- 439 [18] Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization
440 with Some Invalid Instruments Using a Weighted Median Estimator. *Genet Epidemiol.* 2016
441 May;40(4):304-314.

- 442 [19] Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the
443 suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger
444 regression: the role of the I² statistic. *Int J Epidemiol*. 2016 Dec 1;45(6):1961-1974.
- 445 [20] Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization
446 via the zero modal pleiotropy assumption. *Int J Epidemiol*. 2017 Dec 1;46(6):1985-1998.
- 447 [21] Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal
448 relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet*.
449 2018 May;50(5):693-698.
- 450 [22] Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, Frayling TM, Hirschhorn J, Yang
451 J, Visscher PM; GIANT Consortium. Meta-analysis of genome-wide association studies for height and
452 body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet*. 2018 Oct
453 15;27(20):3641-3649.
- 454 [23] Jian H, Xu Q, Wang X, Liu Y, Miao S, Li Y, Mou T, Dong X, Zou X. Amino Acid and Fatty Acid Metabolism
455 Disorders Trigger Oxidative Stress and Inflammatory Response in Excessive Dietary Valine-Induced
456 MASLD of Laying Hens. *Front Nutr*. 2022 Apr 12;9:849767.
- 457 [24] Sunny NE, Kalavalapalli S, Bril F, Garrett TJ, Nautiyal M, Mathew JT, Williams CM, Cusi K. Cross-talk
458 between branched-chain amino acids and hepatic mitochondria is compromised in nonalcoholic fatty
459 liver disease. *Am J Physiol Endocrinol Metab*. 2015 Aug 15;309(4):E311-9.
- 460 [25] Zarfeshani A, Ngo S, Sheppard AM. Leucine alters hepatic glucose/lipid homeostasis via the
461 myostatin-AMP-activated protein kinase pathway - potential implications for nonalcoholic fatty liver
462 disease. *Clin Epigenetics*. 2014 Nov 18;6(1):27.
- 463 [26] He XD, Gong W, Zhang JN, Nie J, Yao CF, Guo FS, Lin Y, Wu XH, Li F, Li J, Sun WC, Wang ED, An YP, Tang
464 HR, Yan GQ, Yang PY, Wei Y, Mao YZ, Lin PC, Zhao JY, Xu Y, Xu W, Zhao SM. Sensing and Transmitting
465 Intracellular Amino Acid Signals through Reversible Lysine Aminoacylations. *Cell Metab*. 2018 Jan
466 9;27(1):151-166.e6.
- 467 [27] Demirel M, Köktaşoğlu F, Özkan E, Dulun Ağaç H, Gül AZ, Sharifov R, Sarıkaya U, Başaranoğlu M,
468 Selek Ş. Mass spectrometry-based untargeted metabolomics study of non-obese individuals with non-
469 alcoholic fatty liver disease. *Scand J Gastroenterol*. 2023 Jul-Dec;58(11):1344-1350.

- 470 [28] Masarone M, Troisi J, Aglitti A, Torre P, Colucci A, Dallio M, Federico A, Balsano C, Persico M.
471 Untargeted metabolomics as a diagnostic tool in NAFLD: discrimination of steatosis, steatohepatitis and
472 cirrhosis. *Metabolomics*. 2021 Jan 16;17(2):12.
- 473 [29] de Mello VD, Sehgal R, Männistö V, Klåvus A, Nilsson E, Perfilyev A, Kaminska D, Miao Z, Pajukanta
474 P, Ling C, Hanhineva K, Pihlajamäki J. Serum aromatic and branched-chain amino acids associated with
475 NASH demonstrate divergent associations with serum lipids. *Liver Int*. 2021 Apr;41(4):754-763.
- 476 [30] Liang KH, Cheng ML, Lo CJ, Lin YH, Lai MW, Lin WR, Yeh CT. Plasma phenylalanine and glutamine
477 concentrations correlate with subsequent hepatocellular carcinoma occurrence in liver cirrhosis patients:
478 an exploratory study. *Sci Rep*. 2020 Jul 2;10(1):10926.
- 479 [31] Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian
480 randomization. *Genet Epidemiol*. 2016 Nov;40(7):597-608.
- 481
- 482



Assumption:

- ① Relevance: G is associated with the X;
- ② Independence: G is not related to any confounding factors of the exposure-outcome association
- ③ Exclusion restriction: G does not affect Y except through its potential effect on the X.

G,genetic variant ; X,exposure ; Y,outcome

Figure 1 Assumptions of the bi-directional Mendelian randomization study design.

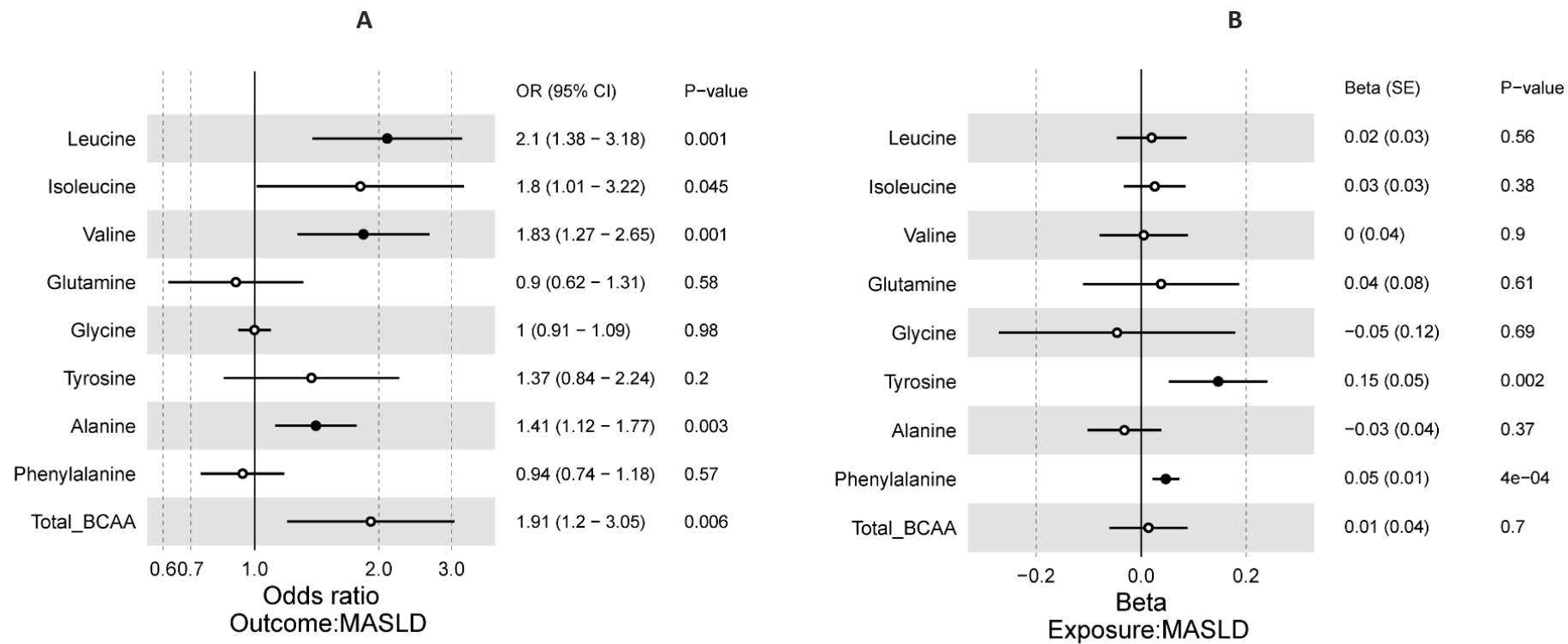


Figure 2 Forest p lots showing the effects of amino acids on MASLD (A) and MASLD on amino acids (B) using the inverse variance-weighted method. CI: Confidence interval. SE: Standard error. Solid dots indicate significance with $p < 6.25 \times 10^{-3}$. Hollow circles indicate non-significance with $p \geq 6.25 \times 10^{-3}$.

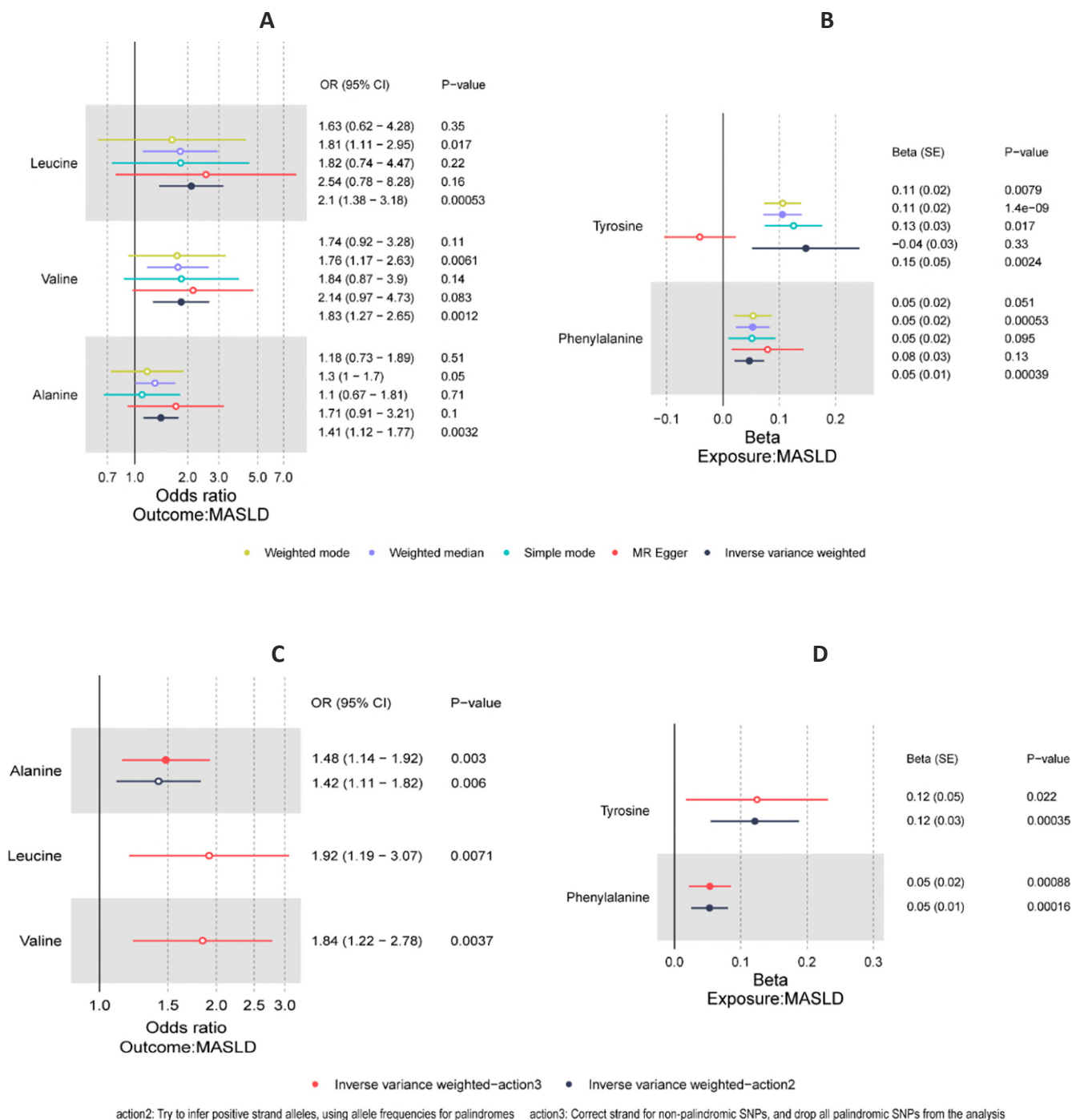


Figure 3 Forest plot showing the effects of leucine, valine, alanine on MASLD and MASLD on tyrosine and phenylalanine. (A-B) Using different Mendelian randomization analyses; (C-D) Using the inverse variance-weighted method with stricter conditions. CI: Confidence interval. SE: Standard error. Solid dots indicate significance with $p < 6.25 \times 10^{-3}$. Hollow circles indicate non-significance with $p \geq 6.25 \times 10^{-3}$.

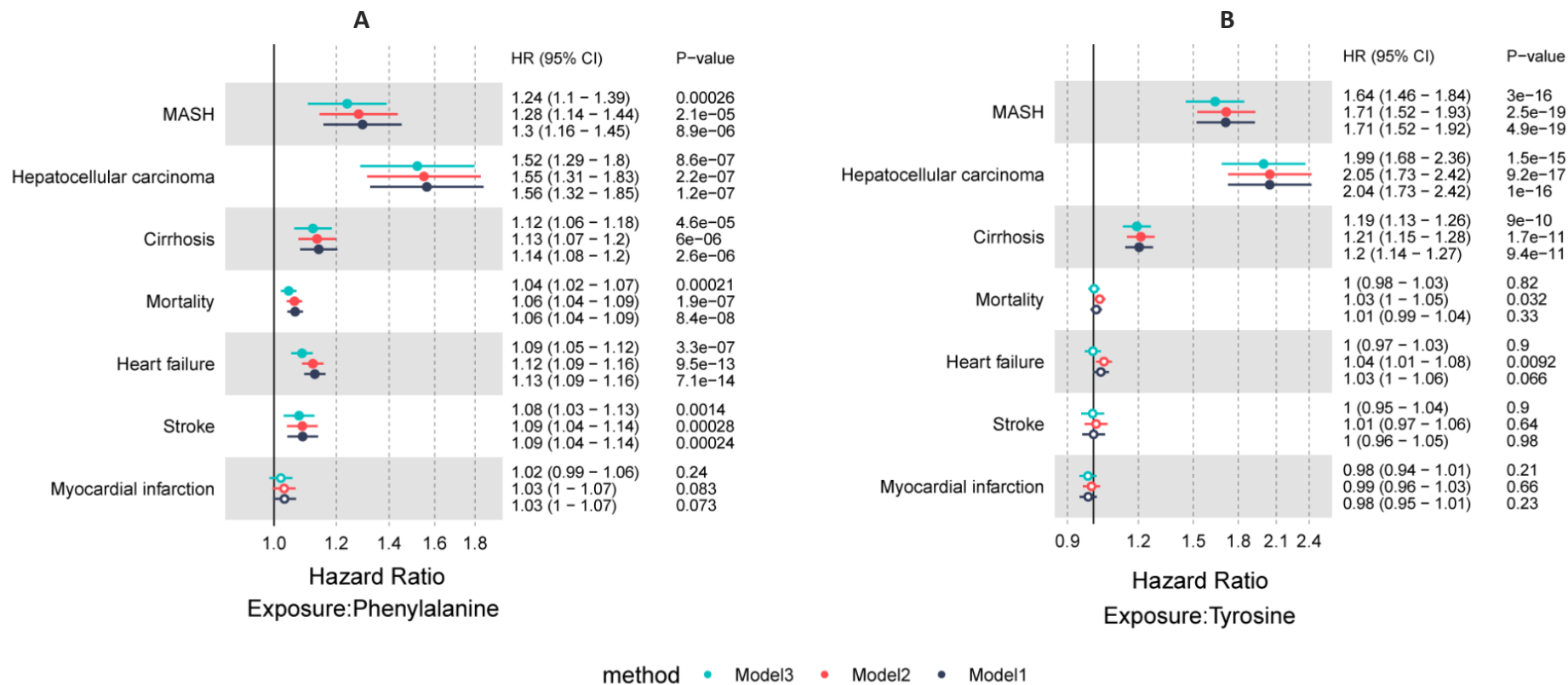


Figure 4 Forest plot showing the associations of phenylalanine (A) and tyrosine (B) with incidence of major outcomes in MASLD patients. CI: Confidence interval. Solid dots indicate significance with $p < 7.1 \times 10^{-3}$. Hollow circles indicate non-significance with $p \geq 7.1 \times 10^{-3}$.

Table 1 Characteristics of study population

MASLD Variables	Control (n = 128,102)	MASLD (n = 72,676)	p-value
Age (years), Mean ± SD	56.21 ± 8.26	57.43 ± 7.86	< 1×10 ⁻⁴
Male, n (%)	41794 (32.63)	42370 (58.3)	< 1×10 ⁻⁴
BMI (kg/m ²), Mean ± SD	24.95 ± 2.88	31.8 ± 4.58	< 1×10 ⁻⁴
Waist circumference (cm), Mean ± SD	82.35 ± 8.93	102.58 ± 10.19	< 1×10 ⁻⁴
Smoking status, n (%)			
Never	81259 (63.43)	38536 (53.02)	
Previous	36400 (28.41)	27040 (37.21)	< 1×10 ⁻⁴
Current	10443 (8.15)	7100 (9.77)	
Pack years [†] , Median (IQR)	0.03 (0.03, 0.03)	0.03 (0.03, 11.5)	< 1×10 ⁻⁴
Alcohol grams per week [‡] , Median (IQR)	44.8 (4, 91.2)	38.4 (4, 101.6)	0.029
Education, n (%)			
College or University degree	45007 (35.13)	18946 (26.07)	
A levels AS levels or equivalent	14645 (11.43)	7251 (9.98)	
CSEs or equivalent	6582 (5.14)	4308 (5.93)	
NVQ or HND or HNC or equivalent	7173 (5.6)	6049 (8.32)	< 1×10 ⁻⁴
O levels GCSEs or equivalent	27987 (21.85)	15359 (21.13)	
Other professional qualifications	6720 (5.25)	4151 (5.71)	
None	19988 (15.6)	16612 (22.86)	
Physical activity, n (%)			
High	55235 (43.12)	24596 (33.84)	
Low	20541 (16.03)	18383 (25.29)	< 1×10 ⁻⁴
Moderate	52326 (40.85)	29697 (40.86)	
Glucose (mmol/L), Mean ± SD	4.95 ± 0.79	5.32 ± 1.31	< 1×10 ⁻⁴
HbA1c (mmol/mol), Mean ± SD	35.06 ± 4.67	38.24 ± 7.66	< 1×10 ⁻⁴
Anti-diabetes medications, n (%)	2336 (1.82)	5898 (8.12)	< 1×10 ⁻⁴
T2D, n (%)	3161 (2.47)	9014 (12.4)	< 1×10 ⁻⁴
Systolic blood pressure (mmHg), Mean ± SD	134.44 ± 18.67	141.49 ± 17.56	< 1×10 ⁻⁴
Diastolic blood pressure (mmHg), Mean ± SD	79.63 ± 9.74	85.01 ± 9.75	< 1×10 ⁻⁴
Anti-hypertensives, n (%)	20703 (16.16)	26087 (35.89)	< 1×10 ⁻⁴
Hypertension, n (%)	54640 (44.52)	49516 (70.33)	< 1×10 ⁻⁴
Triglycerides (mmol/L), Median (IQR)	1.22 (0.92, 1.65)	2.13 (1.58, 2.89)	< 1×10 ⁻⁴
HDL-cholesterol (mmol/L), Mean ± SD	1.54 ± 0.36	1.22 ± 0.28	< 1×10 ⁻⁴
Lipid lowering drugs, n (%)	20530 (16.03)	22364 (30.77)	< 1×10 ⁻⁴

AST, Median (IQR)	23.3 (20.2, 27)	25.9 (22.1, 30.9)	< 1×10 ⁻⁴
ALT, Median (IQR)	17.5 (13.97, 22.37)	25.44 (19.43, 34.26)	< 1×10 ⁻⁴
GGT, Median (IQR)	20.5 (15.9, 28)	36.4 (26.4, 54.4)	< 1×10 ⁻⁴
CRP, Median (IQR)	0.98 (0.51, 1.97)	2.24 (1.19, 4.31)	< 1×10 ⁻⁴

[†]: Never smokers were not considered when calculating pack years. [‡]: Never drinkers were not considered when calculating alcohol grams per week. SD: standard deviation; BMI: body mass index; IQR: interquartile range; AS: Advanced Subsidiary; CES: Certificate of Secondary Education; NVQ: National Vocational Qualification; HND: Higher National Diploma; HNC: Higher National Certificates; GCSE: General Certificate of Secondary Education; HbA1c: Haemoglobin A1c; HDL: high-density lipoprotein; AST: Aspartate aminotransferase; ALT: Alanine aminotransferase; GGT: γ -glutamyl transferase; CRP: c-reactive protein.

Table 2 Observational association of plasma amino acids and MASLD

Amino acids	Model 1 (age+sex+fasting time)			Model 2 (Model 1+lifestyle factors)			Model 3 (Model 2+BMI)		
	beta	se	p-value	beta	se	p-value	beta	se	p-value
Valine	0.685	0.006	0	0.692	0.006	0	0.595	0.009	0
Leucine	0.492	0.005	0	0.494	0.005	0	0.411	0.008	0
Isoleucine	0.489	0.005	0	0.487	0.005	0	0.438	0.008	0
Total BCAA	0.616	0.006	0	0.620	0.006	0	0.533	0.009	0
Histidine	-0.037	0.005	2.72×10^{-14}	-0.029	0.005	6.57×10^{-9}	0.034	0.008	7.30×10^{-6}
Glutamine	-0.337	0.005	0	-0.340	0.005	0	-0.242	0.008	1.60×10^{-207}
Glycine	-0.461	0.005	0	-0.458	0.005	0	-0.313	0.008	6.04×10^{-308}
Tyrosine	0.447	0.005	0	0.455	0.005	0	0.300	0.008	2.42×10^{-296}
Alanine	0.300	0.005	0	0.305	0.005	0	0.339	0.008	0
Phenylalanine	0.305	0.005	0	0.297	0.005	0	0.188	0.008	1.40×10^{-128}

Case/control sample size: (72,626/128,102). Lifestyle factors: smoking status, pack years, alcohol grams per week, education, and physical activity.

Table 3 IVW-MR results and sensitivity analysis of the association between BMI (exposure) and plasma amino acids associated to MASLD (outcome)

Outcome	beta	se	p-value	Pleiotropy (p-value)	MR-PRESSO (p-value)
Leucine	0.189	0.017	1.84×10^{-27}	0.446	0.63
Valine	0.232	0.018	1.36×10^{-37}	0.725	0.491
Alanine	0.041	0.018	0.02	0.318	0.768

Table. 4 Multivariable MR results of the association between plasma amino acids (exposure) and MASLD (outcome), mediating by BMI effect

exposure	BMI				Amino acids			
	SNP number	beta	se	p-value	SNP number	beta	se	p-value
Leucine	479	0.340	0.061	2.35×10^{-8}	2	2.904	-	-
Alanine	469	0.442	0.056	5.33×10^{-15}	4	0.581	0.576	0.31
Valine	475	0.295	0.062	1.78×10^{-6}	3	1.278	0.342	1.9×10^{-4}