

1 **Relationship between branched chain amino acids and type 2 diabetes: a bidirectional**  
2 **Mendelian Randomization study**

3

4 Jonathan D. Mosley<sup>1,2\*</sup>, Mingjian Shi<sup>1\*</sup>, David Agamasu<sup>3</sup>, Nataraja Sarma Vaitinadin, PhD<sup>2</sup>,  
5 Venkatesh Murthy<sup>4</sup>, Ravi V. Shah<sup>2</sup>, Minoos Bagheri<sup>2</sup>, Jane F. Ferguson<sup>2</sup>

6 \*Contributed equally.

7

8 1) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN,  
9 USA; 2) Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA;  
10 3) Meharry Medical College, Nashville, Tennessee, USA; 4) Division of Cardiovascular  
11 Medicine, Department of Medicine, University of Michigan, Ann Arbor, MI

12

13 **Address for correspondence:**

14 Jonathan D. Mosley, M.D., Ph.D.

15 Associate Professor of Medicine and Biomedical Informatics

16 Vanderbilt University Medical Center

17 1285 Medical Research Building IV

18 Nashville, TN 37232

19 Telephone: 615-875-2965; Fax: 615-343-4522; E-mail: [jonathan.d.mosley@vumc.org](mailto:jonathan.d.mosley@vumc.org)

20

21 **Keywords:** Branched chain amino acids, type 2 diabetes, Mendelian randomization, insulin  
22 resistance, body mass index

23

24 **ABSTRACT**

25 **Background:** Human genetic studies suggest that the branched chain amino acids (BCAA)  
26 valine, leucine and isoleucine have a causal association with type 2 diabetes. However,  
27 inferences are based on analyses of a limited number of genetic loci associated with BCAAs.  
28 Whether these conclusions are supported when using instrumental variables for BCAAs that  
29 capture a broad set of genetic mechanisms is not known. **Methods:** We constructed and  
30 validated instrumental variables for each BCAA using large well-powered datasets and tested  
31 their association with type 2 diabetes using the two-sample inverse variance weighted (IWV)  
32 Mendelian randomization (MR) approach. Sensitivity analyses were performed to ensure the  
33 accuracy of the findings. Instrumental variables for type 2 diabetes, fasting insulin and body  
34 mass index (BMI) were also tested for associations with BCAA levels. **Results:** There were no  
35 significant associations with diabetes for valine (beta=0.17 change in log-odds per standard  
36 deviation change in valine, [95% CI, -0.28 - 0.62], p=0.45), leucine (beta=0.19 [-0.30 - 0.68]  
37 p=0.45) or isoleucine (beta=0.02 [-0.54 - 0.59], p=0.94). In contrast, type 2 diabetes was  
38 associated with each BCAA (valine: beta=0.08 per standard deviation change in levels per log-  
39 odds change in type 2 diabetes, [0.05 - 0.10], p=1.8x10<sup>-9</sup>), (leucine: beta= 0.06 [0.04 - 0.09],  
40 p=4.5x10<sup>-8</sup>) and isoleucine (beta= 0.06 [0.04 - 0.08], p=2.8x10<sup>-8</sup>). The type 2 diabetes  
41 associations were replicated in an independent population, but not in a second population where  
42 type 2 diabetes cases were removed, highlighting the consistency and specificity of the  
43 association. Similar positive associations were seen for fasting insulin and BMI with the BCAAs.  
44 In multivariable MR analyses, type 2 diabetes and fasting insulin had consistent independent  
45 associations with each BCAA. **Conclusions:** These data suggest that the BCAAs are not  
46 mediators of type 2 diabetes risk but are biomarkers of diabetes and higher insulin.

47 **INTRODUCTION**

48 The branched chain amino acids (BCAA) valine, leucine and isoleucine are essential amino acids  
49 with functional roles that include regulation of lipid, glucose and protein metabolism.<sup>1</sup> It has long  
50 been recognized that BCAA levels in blood are elevated in individuals with impaired insulin  
51 secretion and obesity.<sup>2</sup> These early observations have prompted interest as to whether the  
52 BCAAs may have an etiological role in insulin resistance and metabolic disease, including type 2  
53 diabetes. Human studies suggesting roles for the BCAAs as upstream drivers of metabolic  
54 disease include a prospective study of normoglycemic participants that found individuals with  
55 the highest baseline BCAA levels were at 5-fold increased risk of incident diabetes.<sup>3</sup> However,  
56 findings from epidemiological studies have not been consistent in all populations.<sup>4-8</sup>

57  
58 Support for an etiological role of the BCAAs in the development of type 2 diabetes has also  
59 come from human genetics studies. Genetic methodologies, such as Mendelian randomization,  
60 use single nucleotide polymorphisms (SNPs) as instrumental variables to determine whether an  
61 exposure may have a causal association with an outcome.<sup>9</sup> An advantage of these studies is that  
62 they can be less prone to biases such as reverse causation, which can lead to incorrect inferences  
63 in epidemiological studies.<sup>10</sup> One study focused largely on SNPs flanking the gene *PPMIK*,  
64 which encodes a protein phosphatase which activates the branched chain  $\alpha$ -ketoacid  
65 dehydrogenase complex (BCKDC) that catalyzes the first irreversible step in the BCAA  
66 metabolism.<sup>11</sup> The investigators identified a significant positive association between genetically  
67 predicted BCAA levels and type 2 diabetes risk, suggesting that BCAAs are upstream mediators  
68 of risk.<sup>12</sup> A subsequent study examined 3 genetic instruments near the *PPMIK*, *SLCIA4*, *ASGRI*  
69 genes and did not find an association with Homeostatic Model Assessment for Insulin Resistance

70 (HOMA-IR).<sup>13</sup> Rather they found a polygenic predictor of HOMA-IR was positively associated  
71 with the BCAAs, suggesting that the BCAAs are elevated as a consequence of insulin resistance.  
72 These results were supported by a later study showing a positive association between genetic  
73 predictors of insulin resistance phenotypes and BCAAs. A common unifying interpretation of  
74 these findings is that BCAAs may be the causal down-stream mediators linking insulin resistance  
75 and obesity phenotypes to type 2 diabetes risk.<sup>12,14</sup>

76  
77 A limitation of the prior genetic studies is that they examined a small number of genetic loci  
78 associated with BCAA levels, thereby limiting the strength of evidence for a causal role of the  
79 BCAAs on type 2 diabetes. Recent large-scale genome-wide association studies (GWAS) of  
80 BCAAs have identified considerably more genetic loci associated with the BCAAs, providing an  
81 opportunity to investigate etiological relationships using robust predictors that represent  
82 heterogenous genetic mechanisms modulating BCAA levels.<sup>15</sup> We developed genetic  
83 instruments for the BCAAs and metabolic phenotypes using data derived from contemporary  
84 large-scale GWAS and applied Mendelian randomization approaches to define the relationships  
85 among these phenotypes.

86  
87

## 88 **METHODS**

### 89 **GWAS Summary Statistics data sets**

90 Single nucleotide polymorphisms (SNPs) associated with valine, leucine and isoleucine used in  
91 the primary analyses were derived from GWAS summary statistics generated by a study  
92 examining 115,082 UK Biobank European Ancestry participants with NMR metabolomics from

93 Nightingale Health (biomarker quantification version 2020).<sup>15</sup> Prior to GWAS, metabolite levels  
94 were inverse rank-normalized and set to a standard deviation of 1. Data were downloaded from  
95 the NHGRI-EBI GWAS Catalog (GCST90092891, GCST90092843, GCST90092995).<sup>16</sup> GWAS  
96 summary statistics used for validation of SNP genetic instruments were obtained from the  
97 Metabolic Syndrome in Men (METSIM) consortium, a study of 6136 Finnish men without  
98 prevalent or incident type 2 diabetes and with metabolites measured using the relative  
99 quantitative liquid chromatography–tandem mass spectrometry (LC-MS/MS) Metabolon  
100 DiscoveryHD4 platform.<sup>17</sup> Metabolite level residuals were inverse-normalized and summary  
101 statistics were downloaded from <https://pheweb.org/metsim-metab/>. Summary statistics were  
102 also obtained a study of 7,824 European ancestry participants from the TwinsUK and  
103 Cooperative Health Research in the Region of Augsburg (KORA) cohort who underwent  
104 metabolite acquisition by Metabolon using GC-MS or UPLC-MS/MS. Summary statistics based  
105 on log<sub>10</sub> transformed metabolite levels were obtained from [http://metabolomics.helmholtz-](http://metabolomics.helmholtz-muenchen.de/gwas/)  
106 [muenchen.de/gwas/](http://metabolomics.helmholtz-muenchen.de/gwas/).<sup>18</sup>  
107  
108 Summary statistics for type 2 diabetes from a study of 898,130 cases and controls of European  
109 ancestry were obtained from the DIAGRAM consortium ([https://diagram-](https://diagram-consortium.org/downloads.html)  
110 [consortium.org/downloads.html](https://diagram-consortium.org/downloads.html)).<sup>19</sup> Summary statistics were from a subset of the cohort that did  
111 not include the UK Biobank population. Summary statistics of log-transformed fasting insulin  
112 from a study of 151,188 individuals of European ancestry without type 2 diabetes<sup>20</sup> were  
113 obtained from the MAGIC consortium (<https://magicinvestigators.org/downloads/>). Summary  
114 statistics for body mass index (BMI) were from a study of 339,224 adults of European ancestry  
115 and obtained from the GIANT consortium

116 ([https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT\\_consortium\\_data\\_files](https://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files)).<sup>2</sup>

117 <sup>1</sup> None of these data sets included participants from the UK Biobank to ensure there was no  
118 overlap with the UK Biobank metabolite cohort.

119

## 120 **Analysis**

### 121 **Lead SNP identification**

122 Lead SNPs associated with valine, leucine and isoleucine were selected from the UK Biobank  
123 summary statistics using a clumping algorithm that selected a linkage disequilibrium (LD)-  
124 reduced set of SNPs ( $r$ -squared  $< 0.001$ , selection window of 3000 Kb) and a minor allele  
125 frequency (MAF)  $> 1\%$ .

126

### 127 **Annotation**

128 The FUMA webserver (<https://fuma.ctglab.nl/>), with default settings, was used to identify nearby  
129 genes and expression QTLs (based on GTEx v8.0) associated with candidate SNPs.<sup>22</sup> The  
130 DAVID program was used to identify ontologies, cellular locations and pathways associated  
131 with candidate genes.<sup>23,24</sup> These data in conjunction with a manual literature review was used to  
132 identify the likely candidate genes located near associated SNPs. If a single likely candidate gene  
133 could not be identified, all plausible candidate genes are presented. Visualization of the cellular  
134 locations of candidate genes were created with BioRender.com.

135

### 136 **Mendelian Randomization**

137 Mendelian randomization (MR) approaches were used to test associations between genetic  
138 instruments for the BCAAs and outcomes. MR is an analytic framework that uses SNPs

139 associated with diverse pleiotropic mechanisms as instrumental variables to test the association  
140 between an exposure and outcome.<sup>9,10</sup> The validity of the approach relies on the following  
141 assumptions: (1) Relevance: the SNP instrumental variables are associated with the exposures;  
142 (2): Independence: the instrumental variables are not associated with confounders; and (3)  
143 Exclusion restriction: the instrumental variables association with the outcome is only through the  
144 exposure.<sup>25</sup> We demonstrate the relevance assumption by validating that our BCAA genetic  
145 instruments associated with their corresponding BCAAs in an independent population. We  
146 conducted a range of MR sensitivity analyses, as well as multivariable analyses to decrease the  
147 likelihood of violating the other assumptions.  
148  
149 SNP genetic instruments were constructed for each BCAA from the UK Biobank GWAS. Non-  
150 palindromic common (MAF>5%) SNP instruments were selected using a clumping algorithm  
151 and SNPs that were available on each set of summary statistics for the METSIM BCAAs, type 2  
152 diabetes and fasting insulin levels. Two-sample MR approach was used to test the association  
153 between BCAA and outcome (the corresponding BCAA in METSIM and type 2 diabetes) using  
154 inverse-variance weighted regression (IVW).<sup>26</sup> The Cochran's Q statistic and its associated p-  
155 value, was used to assess for horizontal pleiotropy, and  $p < 0.05$  was deemed suggestive of  
156 pleiotropy. In addition, sensitivity analyses were performed using MR-Egger to assess for  
157 horizontal pleiotropy and the weighted median method.<sup>27,28</sup> A leave-one-out analyses was also  
158 performed for the BCAA/type 2 diabetes associations. Associations were tested using the  
159 Mendelian Randomization R package.<sup>29</sup> MR-PRESSO was used to assess for horizontal  
160 pleiotropy (global test) and distortions in IVW estimates (distortion test) with exclusion of SNPs  
161 identified as outliers at  $p < 0.05 / (\# \text{ of genetic instruments})$ .<sup>30</sup>

162 The same MR approaches were used to assess whether instrumental variables for type 2 diabetes,  
163 fasting insulin and BMI were associated with each BCAA. The association between type 2  
164 diabetes and each BCAA was also examined using metabolites measured in the KORA and  
165 METSIM studies.

166 Pairwise multivariable MR (MVMR) was used to test for whether the associations between type  
167 2 diabetes and the BCAAs was independent of fasting insulin and BMI.

168 All statistical tests were two-sided and a nominal  $p < 0.05$  was considered significant.<sup>31</sup>

169

170

## 171 RESULTS

172 We identified independent common SNPs ( $MAF > 1\%$ ) significantly associated ( $p < 5 \times 10^{-8}$ ) with  
173 the BCAAs from GWAS performed on 115,082 European Ancestry UK Biobank participants.

174 There were 18, 15 and 8 associated SNPs for valine, leucine and isoleucine, respectively (**Table**

175 **1** and **Supplemental Table 1**). Five SNPs were located near genes examined in prior BCAA

176 association studies (*GCKR*, *PPMIK*, *DDX19A*, *SLC1A4* and *ASGR1* [which is near

177 *SLC2A4*]).<sup>12,13</sup> Two associations were near gene loci that have direct roles in BCAA metabolism

178 (*PPMIK*, *BCAT2*), while others had known functionality related to intermediate metabolism

179 (*SLC1A4*, *PRODH*, *GLS2*, *GLUD1*, *PCCB*), insulin signaling (*IRS1*, *GRB14*, *GCKR*) and

180 glucose regulation (*GCKR*, *SLC2A4*) (**Figure 1**). *HNF4G* and *RLN3* have previously been

181 associated with adiposity traits. We did not observe associations previously reported near

182 *TRMT61A* and *CBLN1* in the UK biobank or METSIM cohorts (**Supplemental Table 2**).<sup>12</sup>

183 We constructed genetic instruments for each BCAA in the UK Biobank set (**Supplemental**

184 **Table 3**). To validate these predictors, we tested their associations with corresponding BCAAs



185 measured in the METSIM data set. Effect sizes for each SNP instrument were generally linearly  
186 related for each BCAA between the studies (**Figure 2**), and MR analyses demonstrated  
187 significant ( $p < 0.05$ ), positive associations for each BCAA, without heterogeneity (Cochran's Q  
188 statistic  $p > 0.05$ ) (**Figure 2, Supplemental Table 4**).

189  
190 There was no genetic association between the BCAA predictors and type 2 diabetes by IVW for  
191 valine (beta=0.17 change in log-odds per standard deviation change in valine, [95% CI, -0.28 -  
192 0.62],  $p=0.45$ ), leucine (beta=0.19 [-0.30 - 0.68]  $p=0.45$ ) or isoleucine (beta=0.02 [-0.54 - 0.59],  
193  $p=0.94$ ) by the IVW method (**Figure 3**). Similar results were seen in sensitivity analyses using  
194 the leave-one-out, MR-Egger and weighted median methods (**Supplementary Figure 1,**  
195 **Supplemental Table 5**). There was significant heterogeneity (Cochran's Q statistic  $p < 0.05$ ) for  
196 each association, but the MR-Egger intercept values were not significant ( $p > 0.05$ ), and outlier  
197 removal by the MR-PRESSO method did not identify significant distortion ( $p > 0.05$ ) in the effect  
198 estimates (**Supplemental Table 5**).

199  
200 We next constructed instrumental variables for type 2 diabetes ( $n=95$  SNPs) to determine  
201 whether the type 2 diabetes was associated with altered BCAA levels. There were significant  
202 positive associations with each BCAA (valine: beta=0.08 per standard deviation change in levels  
203 per log-odds change in type 2 diabetes, [95% CI: 0.05 - 0.10],  $p=1.8 \times 10^{-9}$ ), (leucine: beta= 0.06  
204 [0.04 - 0.09],  $p=4.5 \times 10^{-8}$ ) and isoleucine (beta= 0.06 [0.04 - 0.08],  $p=2.8 \times 10^{-8}$ ) (**Figure 5,**  
205 **Supplemental Table 6**). Sensitivity analyses showed consistent results. Similar significant  
206 positive associations were seen when testing associations with with BCAA levels measured in  
207 the KORA study (**Supplementary Figure 2, Supplemental Table 7**). We tested associations

208 with BCAA levels in METSIM. In this set, all participants with prevalent or incident type 2  
209 diabetes were excluded from the BCAA GWAS, which would be expected to attenuate or  
210 eliminate an association with type 2 diabetes if the association was due to this disease. As  
211 expected, while the directions for all associations were consistent with those observed in the UK  
212 Biobank and KORA sets, they were not significant for valine and leucine and nominally  
213 significant for isoleucine ( $p=0.03$ ) (**Supplementary Figure 3, Supplemental Table 8**).

214  
215 The SNPs associated with BCAA levels are associated with insulin regulation and adiposity,  
216 suggesting that the BCAAs may be up-stream of these traits. To test this hypothesis, we  
217 constructed instrumental variables for fasting insulin levels ( $n=34$  SNPs) and BMI ( $n=69$  SNPs).  
218 Similar to type 2 diabetes, there were significant positive associations with each phenotype and  
219 each BCAA (**Supplemental Figure 4, Supplemental Table 9**). In sensitivity analyses the MR-  
220 Egger results for fasting insulin demonstrated discordant results with wide confidence intervals  
221 (**Supplemental Table 9**). However, the intercept  $p$ -values were not significant. MR-PRESSO  
222 analysis did not show significant distortion ( $p>0.05$ ) after removing outliers. It has previously  
223 been observed the MR-Egger can produce unreliable discordant estimates with wide confidence  
224 intervals in the presence of SNPs with a direct effects on the outcome.<sup>28,32</sup> This was the case  
225 here, as, for instance, excluding 2 outlying SNPs identified by MR-PRESSO brought the MR-  
226 Egger estimate (0.61) in line with the IVW (0.60) and weighted median (0.70) estimates for  
227 leucine. In sum, these results suggest that BCAA levels are higher with higher type 2 diabetes  
228 risk, fasting insulin levels and BMI. Multivariable MR was used to determine whether the  
229 associations with type 2 diabetes was independent of BMI and levels. Type 2 diabetes was  
230 associated with each BCAA, independent of BMI and fasting insulin levels (**Figure 5**,

231 **Supplemental Table 8**). However, associations between BMI and the BCAAs were not  
232 significant ( $p>0.05$ ) after adjustment for type 2 diabetes.

233

234

## 235 **DISCUSSION**

236 We constructed and validated genetic instruments for the BCAAs valine, leucine and isoleucine  
237 and tested whether they were associated with type 2 diabetes using standard MR approaches.

238 While the genetic instruments robustly associated with positive control phenotypes, they did not  
239 associate with type 2 diabetes. This contrasts with a prior MR study which demonstrated  
240 associations for each BCAA.<sup>12</sup> In line with prior studies, we found that fasting insulin and BMI  
241 were positively associated with BCAA levels.<sup>13,33</sup> In addition, we found that genetic instruments  
242 for type 2 diabetes were also robustly associated with BCAA levels, suggesting that the BCAAs  
243 are biomarkers, rather than causal mediators of type 2 diabetes. Multivariable MR suggested that  
244 the associations with BMI are likely mediated through increasing type 2 diabetes risk.

245

246 We did not replicate findings from a prior study that observed associations between genetic  
247 instruments of the BCAAs and type 2 diabetes.<sup>12</sup> There are important differences between that  
248 study and this one. The BCAA predictors in that study were heavily weighted with SNPs in high  
249 linkage-disequilibrium around the *PPMK1* gene, whose gene product regulates catabolism of  
250 each of the BCAAs. Their valine and leucine predictors only used SNPs around that gene locus,  
251 and thus were only testing for an association based on a single genetic mechanism. Favoring an  
252 association was the fact that there was a suggestive association around the *PPMK1* gene in their  
253 type 2 diabetes GWAS and the high representation of SNPs in high LD in this region. We did not

254 observe a type 2 diabetes association at the *PPMK1* locus in the DIAGRAM Consortium GWAS  
255 data set. Their isoleucine predictor additionally included SNPs around the *TRMT61A*, *CBLN1*  
256 and *DDX19A*, but not *GCKR*. We did not observe associations and found directional  
257 inconsistency between the BCAAs and their lead SNPs around the *TRMT61A* and *CBLN1* genes  
258 in the UK Biobank and METSIM cohorts. Thus, we could not replicate their findings due to  
259 these non-replicating SNP associations related to isoleucine and type 2 diabetes.

260  
261  
262 We used a larger set of SNPs measuring diverse genetic mechanisms associated with levels of  
263 BCAAs in plasma than prior studies. In general, MR results are most robust when the genetic  
264 instruments measure a broad set of independent genetic mechanisms associated with an  
265 exposure, as this decreases the likelihood of spurious associations that can result from examining  
266 a small number of genetic loci.<sup>26</sup> Attesting to the validity of the BCAA genetic instruments, we  
267 found that the SNP effect sizes corresponded to those measured in an independent cohort. When  
268 testing for the type 2 diabetes association, we employed a two-sample design to reduce the  
269 chance of identifying associations based on reverse causality and we performed a number of  
270 sensitivity analyses to ensure that outlying SNPs were not obscuring an association. Thus, we  
271 believe these findings more accurately describe the genetic relationship between the BCAAs and  
272 type 2 diabetes.

273  
274 The SNPs associated with the BCAAs highlight a number of genetic mechanisms influencing  
275 circulating BCAA levels, many of which have functions within the mitochondria. In addition to  
276 *PPMK1*, the protein product of *BCAT2* transaminates the BCAAs to  $\alpha$ -ketoacids while

277 converting  $\alpha$ -ketoglutarate ( $\alpha$ KG) to glutamate.<sup>34</sup> The gene products of *GLUDI*, *PRODH* and  
278 *GLS2* also impact glutamate metabolism.<sup>35</sup> *PCCB* contributes to BCAA breakdown by  
279 catabolism of propionyl-CoA.<sup>36</sup> Several genes are related to glucose regulation, including *GCKR*  
280 which inhibits glucose metabolism by glucokinase, *SLC2A4* (*GLUT4*) which allows glucose  
281 entry into cells in response to insulin signaling, and glucose signaling including *MLXIP* which  
282 modulates transcription in response to intracellular glucose levels.<sup>37</sup> *IRS1* and *GRB14* both  
283 regulate signaling through the insulin receptor.<sup>38</sup> There were also genes associated to adiposity  
284 including *HNF4G*, a transcription factor found to be associated with BMI in GWAS studies and  
285 *RNL3*, which may have a role in appetite regulation.<sup>39-41</sup> There were also SNPs located at loci  
286 that identified genomic regions where the link to BCAAs was not clear including SNPs near  
287 *MOK*, *DDX19A-DT*, *SUOX*<sup>42</sup> and *DHODH*.

288  
289 Collectively, the SNP associations support the hypothesis that BCAA levels are influenced not  
290 only by gene products directly involved in BCAA catabolism, but also physiological processes  
291 related to insulin signaling, glucose regulation and adiposity.<sup>43</sup> Consistently, we found that  
292 genetic instruments related to BMI, fasting insulin levels and type 2 diabetes associated with  
293 each BCAA. The associations were in a positive direction, consistent with well-established  
294 epidemiological observations. For leucine and isoleucine, the BMI association was not  
295 significant after multivariable adjustment, and remained only nominally associated with valine.  
296 These results suggest that the BMI association was most likely related to the influences of  
297 adiposity on insulin resistance and penetrance of type 2 diabetes.<sup>44</sup>

298

299 There are limitations to the current study. False positive associations can arise from the use of  
300 weak genetic instruments or the use of pleiotropic SNPs that associate with a risk factor through  
301 mechanisms unrelated to the exposure. To mitigate this possibility, we used a range of sensitivity  
302 analyses to ensure that associations were consistent across a range of assumptions. These  
303 findings also conflict with some epidemiological and experimental data, which could suggest  
304 that non-genetically determined BCAA levels may contribute to type 2 diabetes and insulin  
305 resistance through mechanisms not captured by these studies. The analyses also relied on  
306 existing data that predominantly represents individuals of European ancestries, which limits  
307 generalization to other ancestry groups.

308

309 In summary, we did not find evidence supporting a causal role for BCAAs on type 2 diabetes  
310 risk. Rather, these data support the hypothesis that the BCAAs levels are increased as a  
311 consequence of diabetes and insulin resistance.

312

313

#### 314 **AUTHOR CONTRIBUTIONS**

315 JDM and MS had full access to all data and takes responsibility for the integrity of the data and  
316 analysis. JDM and JFF provided substantial contributions to the design of the study. MS  
317 performed the primary analysis of the BioVU data. MS and DA made substantial contributions to  
318 the acquisition of data. JDM, MS and JFF wrote the first draft of the manuscript. All coauthors  
319 critically reviewed the manuscript for important intellectual content, provided final approval of  
320 the version to be published, and agree to be accountable for some or all aspects of the work  
321 presented.

322 **DECLARATION OF INTERESTS**

323 None.

324

325 **FUNDING SOURCES**

326 This work was supported by the NIH R01 HL142856 (to JFF and JDM). JFF is also supported by

327 R01 DK117144. MB is supported by T32 HG008341.

328

## 329 REFERENCES

- 330 1. Branched Chain Amino Acids: Beyond Nutrition Metabolism - PubMed [Internet]. [cited  
331 2023 Apr 20]; Available from: <https://pubmed.ncbi.nlm.nih.gov/29570613/>
- 332 2. Felig P, Marliss E, Cahill GF. Plasma amino acid levels and insulin secretion in obesity. *N*  
333 *Engl J Med.* 1969;281:811–816.
- 334 3. Wang TJ, Larson MG, Vasani RS, Cheng S, Rhee EP, McCabe E, Lewis GD, Fox CS,  
335 Jacques PF, Fernandez C, et al. Metabolite profiles and the risk of developing diabetes. *Nat.*  
336 *Med.* 2011;17:448–453.
- 337 4. Floegel A, Stefan N, Yu Z, Mühlenbruch K, Drogan D, Joost H-G, Fritsche A, Häring H-U,  
338 Hrabě de Angelis M, Peters A, et al. Identification of serum metabolites associated with  
339 risk of type 2 diabetes using a targeted metabolomic approach. *Diabetes.* 2013;62:639–648.
- 340 5. McCormack SE, Shaham O, McCarthy MA, Deik AA, Wang TJ, Gerszten RE, Clish CB,  
341 Mootha VK, Grinspoon SK, Fleischman A. Circulating branched-chain amino acid  
342 concentrations are associated with obesity and future insulin resistance in children and  
343 adolescents. *Pediatr Obes.* 2013;8:52–61.
- 344 6. Merino J, Leong A, Liu C-T, Porneala B, Walford GA, von Grotthuss M, Wang TJ,  
345 Flannick J, Dupuis J, Levy D, et al. Metabolomics insights into early type 2 diabetes  
346 pathogenesis and detection in individuals with normal fasting glucose. *Diabetologia.*  
347 2018;61:1315–1324.
- 348 7. Hamley S, Kloosterman D, Duthie T, Dalla Man C, Visentin R, Mason SA, Ang T,  
349 Selathurai A, Kaur G, Morales-Scholz MG, et al. Mechanisms of hyperinsulinaemia in  
350 apparently healthy non-obese young adults: role of insulin secretion, clearance and action  
351 and associations with plasma amino acids. *Diabetologia.* 2019;62:2310–2324.
- 352 8. Julkunen H, Cichońska A, Tiainen M, Koskela H, Nybo K, Mäkelä V, Nokso-Koivisto J,  
353 Kristiansson K, Perola M, Salomaa V, et al. Atlas of plasma NMR biomarkers for health  
354 and disease in 118,461 individuals from the UK Biobank. *Nat Commun.* 2023;14:604.
- 355 9. Lawlor DA, Harbord RM, Sterne JAC, Timpson N, Davey Smith G. Mendelian  
356 randomization: using genes as instruments for making causal inferences in epidemiology.  
357 *Stat Med.* 2008;27:1133–1163.
- 358 10. Ebrahim S, Davey Smith G. Mendelian randomization: can genetic epidemiology help  
359 redress the failures of observational epidemiology? *Hum. Genet.* 2008;123:15–33.
- 360 11. Biswas D, Duffley L, Pulinilkunnil T. Role of branched-chain amino acid-catabolizing  
361 enzymes in intertissue signaling, metabolic remodeling, and energy homeostasis. *FASEB J.*  
362 2019;33:8711–8731.
- 363 12. Lotta LA, Scott RA, Sharp SJ, Burgess S, Luan J, Tillin T, Schmidt AF, Imamura F,  
364 Stewart ID, Perry JRB, et al. Genetic Predisposition to an Impaired Metabolism of the



- 365 Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian Randomisation  
366 Analysis. *PLoS Med.* 2016;13:e1002179.
- 367 13. Mahendran Y, Jonsson A, Have CT, Allin KH, Witte DR, Jørgensen ME, Grarup N,  
368 Pedersen O, Kilpeläinen TO, Hansen T. Genetic evidence of a causal effect of insulin  
369 resistance on branched-chain amino acid levels. *Diabetologia.* 2017;60:873–878.
- 370 14. Neinast M, Murashige D, Arany Z. Branched Chain Amino Acids. *Annu Rev Physiol.*  
371 2019;81:139–164.
- 372 15. Richardson TG, Leyden GM, Wang Q, Bell JA, Elsworth B, Davey Smith G, Holmes MV.  
373 Characterising metabolomic signatures of lipid-modifying therapies through drug target  
374 mendelian randomisation. *PLoS Biol.* 2022;20:e3001547.
- 375 16. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon  
376 A, Morales J, Mountjoy E, Sollis E, et al. The NHGRI-EBI GWAS Catalog of published  
377 genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic  
378 Acids Res.* 2018;
- 379 17. Yin X, Chan LS, Bose D, Jackson AU, VandeHaar P, Locke AE, Fuchsberger C, Stringham  
380 HM, Welch R, Yu K, et al. Genome-wide association studies of metabolites in Finnish men  
381 identify disease-relevant loci. *Nat Commun.* 2022;13:1644.
- 382 18. Shin S-Y, Fauman EB, Petersen A-K, Krumsiek J, Santos R, Huang J, Arnold M, Erte I,  
383 Forgetta V, Yang T-P, et al. An atlas of genetic influences on human blood metabolites.  
384 *Nat. Genet.* 2014;46:543–550.
- 385 19. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, Payne AJ,  
386 Steinthorsdottir V, Scott RA, Grarup N, et al. Fine-mapping type 2 diabetes loci to single-  
387 variant resolution using high-density imputation and islet-specific epigenome maps. *Nat.  
388 Genet.* 2018;50:1505–1513.
- 389 20. Lagou V, Mägi R, Hottenga J-J, Grallert H, Perry JRB, Bouatia-Naji N, Marullo L, Rybin  
390 D, Jansen R, Min JL, et al. Sex-dimorphic genetic effects and novel loci for fasting glucose  
391 and insulin variability. *Nat Commun.* 2021;12:24.
- 392 21. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S,  
393 Buchkovich ML, Yang J, et al. Genetic studies of body mass index yield new insights for  
394 obesity biology. *Nature.* 2015;518:197–206.
- 395 22. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and  
396 annotation of genetic associations with FUMA. *Nat Commun.* 2017;8:1826.
- 397 23. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the  
398 comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009;37:1–13.
- 399 24. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene  
400 lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4:44–57.

- 401 25. de Leeuw C, Savage J, Bucur IG, Heskes T, Posthuma D. Understanding the assumptions  
402 underlying Mendelian randomization. *Eur J Hum Genet.* 2022;30:653–660.
- 403 26. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple  
404 genetic variants using summarized data. *Genet. Epidemiol.* 2013;37:658–665.
- 405 27. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian  
406 Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet  
407 Epidemiol.* 2016;40:304–314.
- 408 28. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the  
409 MR-Egger method. *Eur J Epidemiol.* 2017;32:377–389.
- 410 29. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing  
411 Mendelian randomization analyses using summarized data. *Int J Epidemiol.* 2017;46:1734–  
412 1739.
- 413 30. Verbanck M, Chen C-Y, Neale B, Do R. Detection of widespread horizontal pleiotropy in  
414 causal relationships inferred from Mendelian randomization between complex traits and  
415 diseases. *Nat Genet.* 2018;50:693–698.
- 416 31. Burgess S, Thompson SG. Multivariable Mendelian randomization: the use of pleiotropic  
417 genetic variants to estimate causal effects. *Am. J. Epidemiol.* 2015;181:251–260.
- 418 32. Gill D, Cameron AC, Burgess S, Li X, Doherty DJ, Karhunen V, Abdul-Rahim AH,  
419 Taylor-Rowan M, Zuber V, Tsao PS, et al. Urate, Blood Pressure, and Cardiovascular  
420 Disease: Evidence From Mendelian Randomization and Meta-Analysis of Clinical Trials.  
421 *Hypertension.* 2021;77:383–392.
- 422 33. Wang Q, Holmes MV, Davey Smith G, Ala-Korpela M. Genetic Support for a Causal Role  
423 of Insulin Resistance on Circulating Branched-Chain Amino Acids and Inflammation.  
424 *Diabetes Care.* 2017;40:1779–1786.
- 425 34. Ichihara A. Isozyme patterns of branched-chain amino acid transaminase during cellular  
426 differentiation and carcinogenesis. *Ann N Y Acad Sci.* 1975;259:347–354.
- 427 35. Hewton KG, Johal AS, Parker SJ. Transporters at the Interface between Cytosolic and  
428 Mitochondrial Amino Acid Metabolism. *Metabolites.* 2021;11:112.
- 429 36. Wongkittichote P, Ah Mew N, Chapman KA. Propionyl-CoA carboxylase - A review. *Mol  
430 Genet Metab.* 2017;122:145–152.
- 431 37. Kaadige MR, Looper RE, Kamalanaadhan S, Ayer DE. Glutamine-dependent anapleurosis  
432 dictates glucose uptake and cell growth by regulating MondoA transcriptional activity. *Proc  
433 Natl Acad Sci U S A.* 2009;106:14878–14883.
- 434 38. Haeusler RA, McGraw TE, Accili D. Biochemical and cellular properties of insulin  
435 receptor signalling. *Nat Rev Mol Cell Biol.* 2018;19:31–44.

- 436 39. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, Lango Allen  
437 H, Lindgren CM, Luan J, Mägi R, et al. Association analyses of 249,796 individuals reveal  
438 18 new loci associated with body mass index. *Nat. Genet.* 2010;42:937–948.
- 439 40. McGowan BMC, Stanley SA, Smith KL, White NE, Connolly MM, Thompson EL,  
440 Gardiner JV, Murphy KG, Ghatei MA, Bloom SR. Central relaxin-3 administration causes  
441 hyperphagia in male Wistar rats. *Endocrinology.* 2005;146:3295–3300.
- 442 41. Ghattas MH, Mehanna ET, Mesbah NM, Abo-Elmatty DM. Relaxin-3 is associated with  
443 metabolic syndrome and its component traits in women. *Clin Biochem.* 2013;46:45–48.
- 444 42. Kappler U, Enemark JH. Sulfite-oxidizing enzymes. *J Biol Inorg Chem.* 2015;20:253–264.
- 445 43. Doestzada M, Zhernakova DV, C L van den Munckhof I, Wang D, Kurilshikov A, Chen L,  
446 Bloks VW, van Faassen M, Rutten JHW, Joosten LAB, et al. Systematic analysis of  
447 relationships between plasma branched-chain amino acid concentrations and  
448 cardiometabolic parameters: an association and Mendelian randomization study. *BMC Med.*  
449 2022;20:485.
- 450 44. Kahn SE, Hull RL, Utzschneider KM. Mechanisms linking obesity to insulin resistance and  
451 type 2 diabetes. *Nature.* 2006;444:840–846.

452

453 **TABLES**

454

455 **Table 1.** Genomic loci and candidate genes associated with circulating valine, leucine or

456 isoleucine levels.

457

<b>Chromosome: Position (Hg37)</b>	<b>Lead SNP(s)</b>	<b>Lead candidate gene(s)<sup>1</sup></b>	<b>Function<sup>2</sup></b>
2:27730940	Val: rs1260326 Leu: rs1260326 Ile: rs1260326	<i>GCKR</i>	Inhibits glucokinase in liver and pancreatic islet cells.
2:65219030	Val: rs2007061	<i>SLCIA4</i>	Sodium-dependent neutral amino acid transporter for alanine, serine, cysteine, and threonine.
2:165501927 - 2:165528876	Val: rs13389219 Leu: rs5835988 Ile: rs5835988	<i>GRB14</i>	Interacts with insulin receptors and inhibits insulin receptor signaling.
2:227108446	Val: rs2943652 Leu: rs2943652	<i>IRS1</i>	Transmits signal between the insulin receptor and phosphoinositide 3-kinase.
3:136328270	Val: rs1471740 Leu: rs1471740	<i>PCCB</i>	Breaks down amino acids including isoleucine and valine.
4:88920985	Val: rs10018448 Leu: rs114796149, rs10018448 Ile: rs114796149	<i>PPMIK</i>	Activates branched chain $\alpha$ -ketoacid dehydrogenase complex that catalyzes the first step in the BCAA metabolism.
8:76443463 - 8:76454025	Leu: rs2977929 Ile: rs2941456	<i>HNF4G</i>	Enables DNA-binding transcription activator activity. Associated with BMI by GWAS.
10:88820592	Val: rs17096421	<i>GLUD1</i>	Mitochondrial matrix enzyme that catalyzes the oxidative deamination of glutamate to alpha-ketoglutarate and ammonia.
12:56393337	Val: rs1081975 Leu: rs1081975	<i>SUOX</i>	Breaks down sulfur-containing amino acids.
12:56861458 - 12:56865056	Val: rs2638315 Leu: rs2638315 Ile: rs7302925	<i>GLS2</i>	Catalyzes the hydrolysis of glutamine to glutamate and ammonia.
12:122607112	Val: rs12368387	<i>MLXIP (MONDOA)</i>	Transcriptional activation of glycolytic target genes in response to cellular glucose levels.
14:102718052	Val: rs2274815	<i>WDR20, MOK, ZNF839</i>	Multiple. Lead SNP intronic in MOK.

16:70270907 - 16:70368909	Val: rs12325419 Leu: rs35452938 Ile: rs12325419	<i>Multiple, including DDX19A-DT and DDX19B</i>	Multiple. Intronic in DDX19A-DT and 3' UTR in DDX19B. DDX19A identified on prior isoleucine GWAS.
16:72149923	Leu: rs9930957	<i>DHODH, TXNL4B, HP, HPR, DHX38, PMFBP1</i>	Multiple.
17:7185779	Val: rs117643180 Leu: rs117643180 Ile: rs117643180	<i>SLC2A4 (GLUT4)</i>	Insulin-regulated facilitative membrane glucose transport into cells.
17:79636653	Val: rs62080209	<i>SLC25A10 (DIC)</i>	Translocates small metabolites (e.g. malate, succinate) across the mitochondrial membrane for metabolic processes including the TCA cycle and fatty acid synthesis.
19:14139004	Leu: rs1982632	<i>RLN3</i>	Insulin-superfamily hormone regulating appetite, food intake, and weight gain.
19:49300431 - 19:49317459	Val: rs35230038, rs117048185 Leu: rs4801776 Ile: rs71179052	<i>BCAT2</i>	Mitochondrial branched chain aminotransferase that catalyzes production of the BCAAs.
22:18915282 - 22:18915347	Val: rs2238732 Leu: rs5747934	<i>PRODH</i>	Mitochondrial protein that catalyzes the first step in proline degradation.

458

459 Footnotes:

460 1. Lead candidate genes shown were selected based on their proximity to lead the lead SNP,

461 eQTL analyses and known relevance of cellular functions to BCAA metabolism.

462 2. Known cellular functions for loci where a single lead candidate gene is identified.

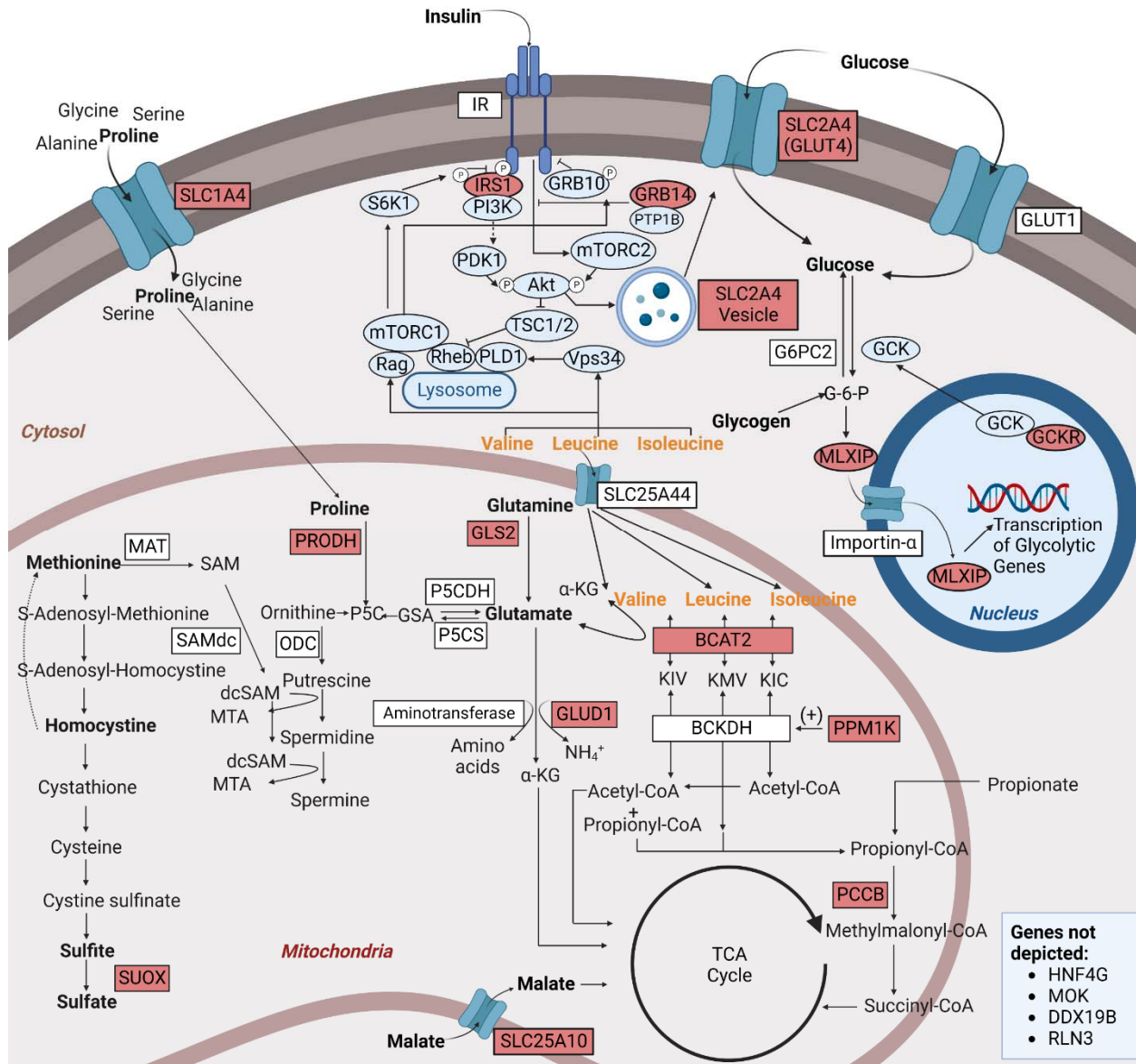
463

464 **FIGURE LEGENDS**

465

466 **Figure 1: Cellular functions of proteins for genes associated with circulating valine, leucine**  
 467 **and isoleucine levels. Genes located near lead SNPs are highlighted in red.**

468

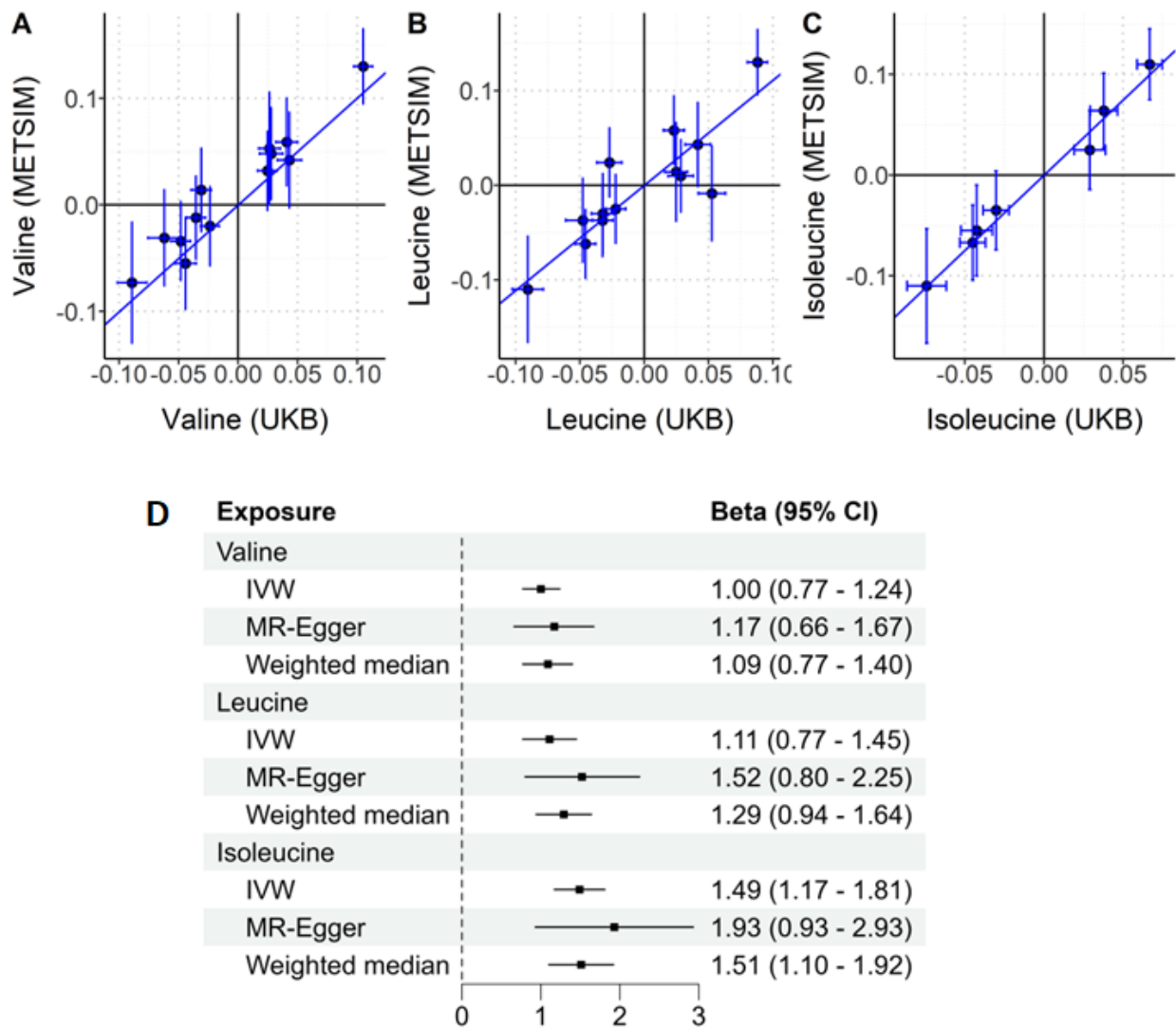


469

470

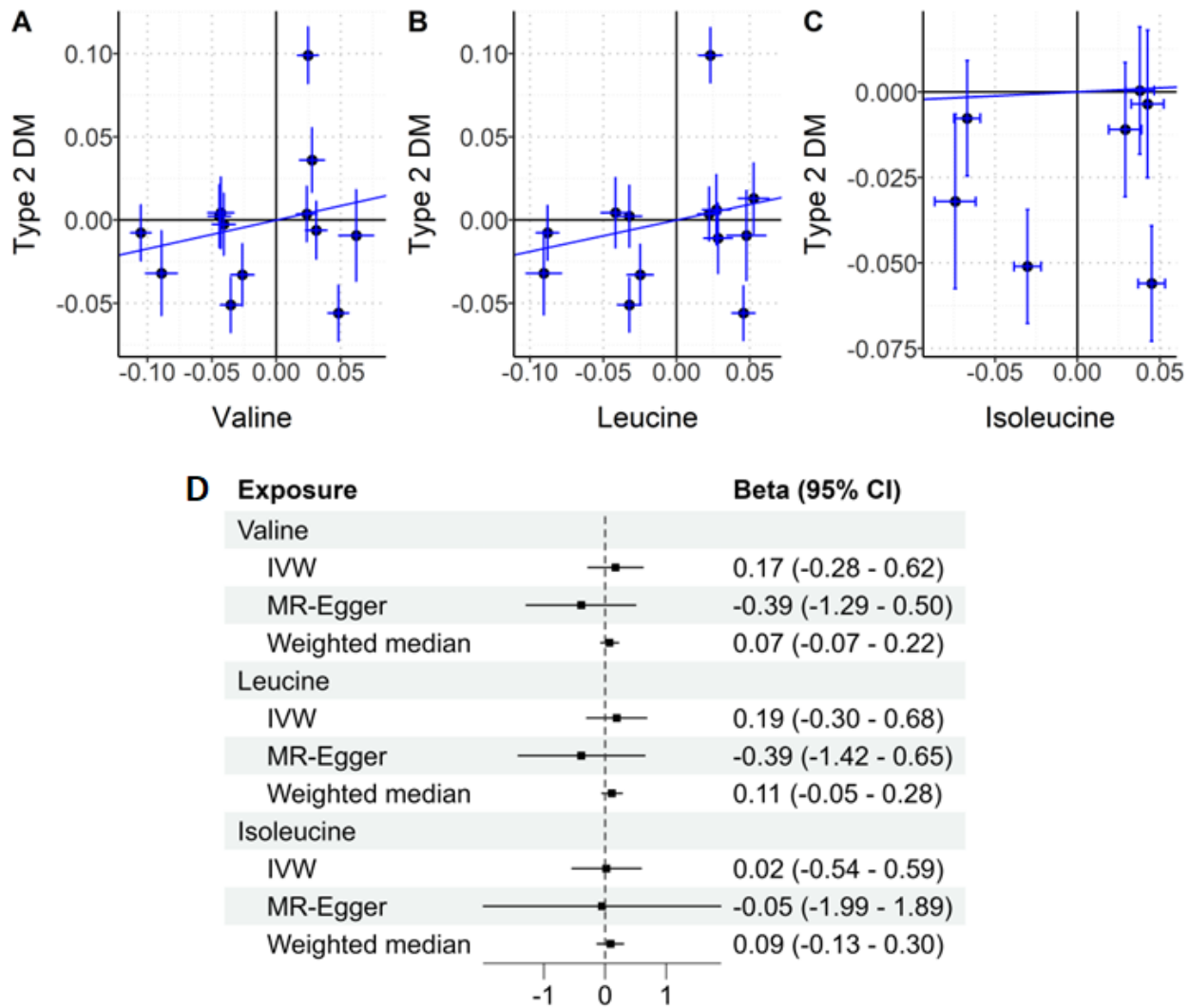
471

472 **Figure 2: Associations between SNP instruments for the BCAAs and validation**  
 473 **phenotypes.** (A-C) Scatter plots showing comparing the GWAS effect sizes between the valine  
 474 (n=13 SNPs), leucine (n=13 SNPs) and isoleucine SNP (n=7 SNPs) genetic instruments  
 475 identified in the UK Biobank cohort and their corresponding amino acid in the METSIM cohort.  
 476 The lines represent the association based on the IVW method. (D) Summary of associations from  
 477 MR analyses testing the association between the genetic instruments in UK Biobank and their  
 478 corresponding amino acid level in METSIM. IVW=inverse variance weighted method.  
 479



480

481 **Figure 3: Associations between SNP instruments for the BCAAs and type 2 diabetes.** (A-C)  
 482 Scatter plots showing comparing the associations between the valine, leucine and isoleucine SNP  
 483 genetic instruments and with type 2 diabetes. The lines represent the association based on the  
 484 IVW method. (D) Summary of associations from MR analyses testing the association between  
 485 the genetic instruments for each BCAA and type 2 diabetes. IVW=inverse variance weighted  
 486 method.  
 487



488



489 **Figure 4: Associations between SNP instruments for metabolic phenotypes and BCAAs. (A-**

490 C) Scatter plots showing of the associations between genetic instruments for type 2 diabetes

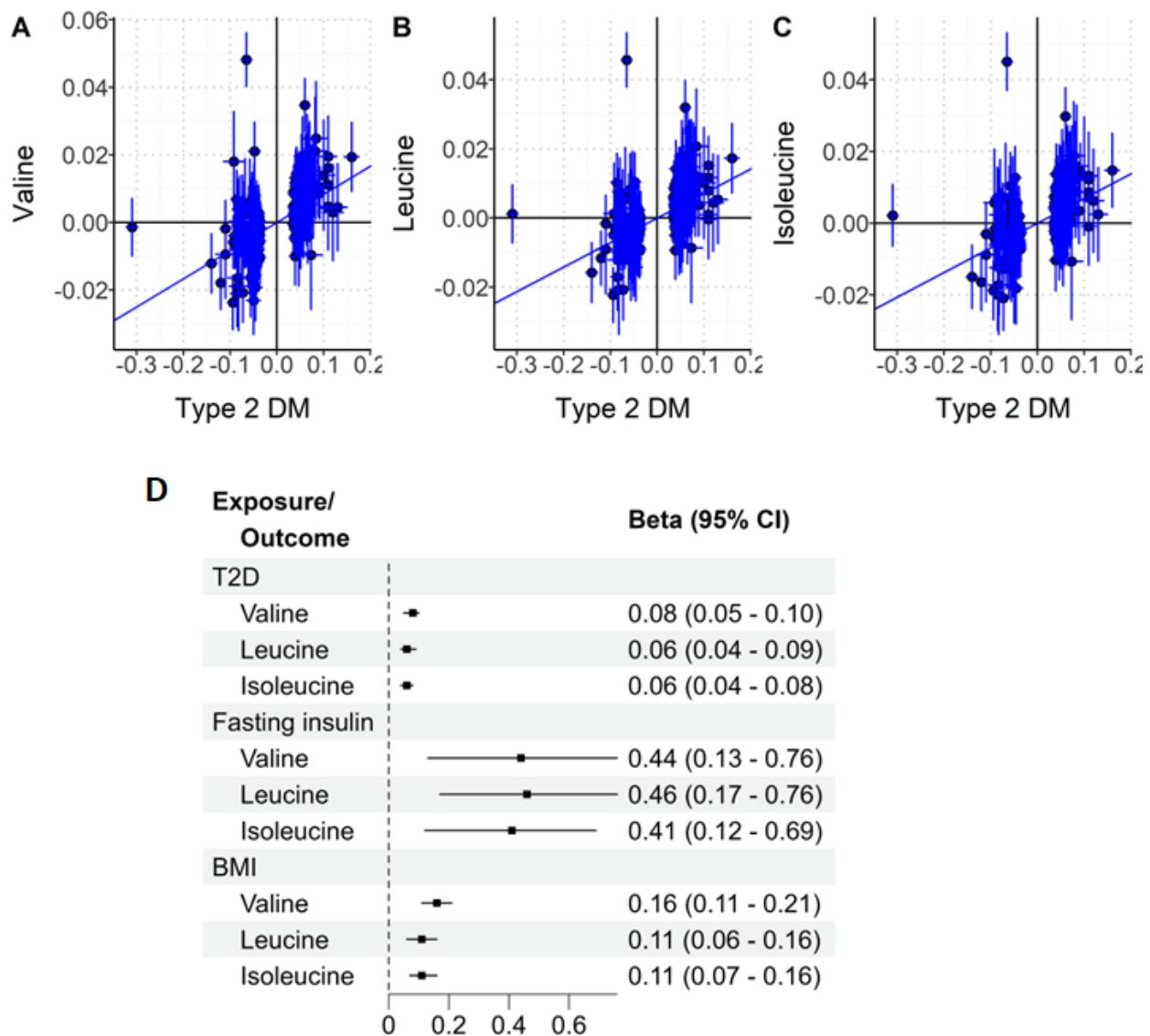
491 (n=95 SNPs) and valine, leucine and isoleucine. The lines represent the association based on the

492 IVW method. (D) Forest plot summarizing the associations by the inverse variance weighted

493 (IVW) method between genetic instruments for type 2 diabetes, fasting insulin (n=34 SNPs) and

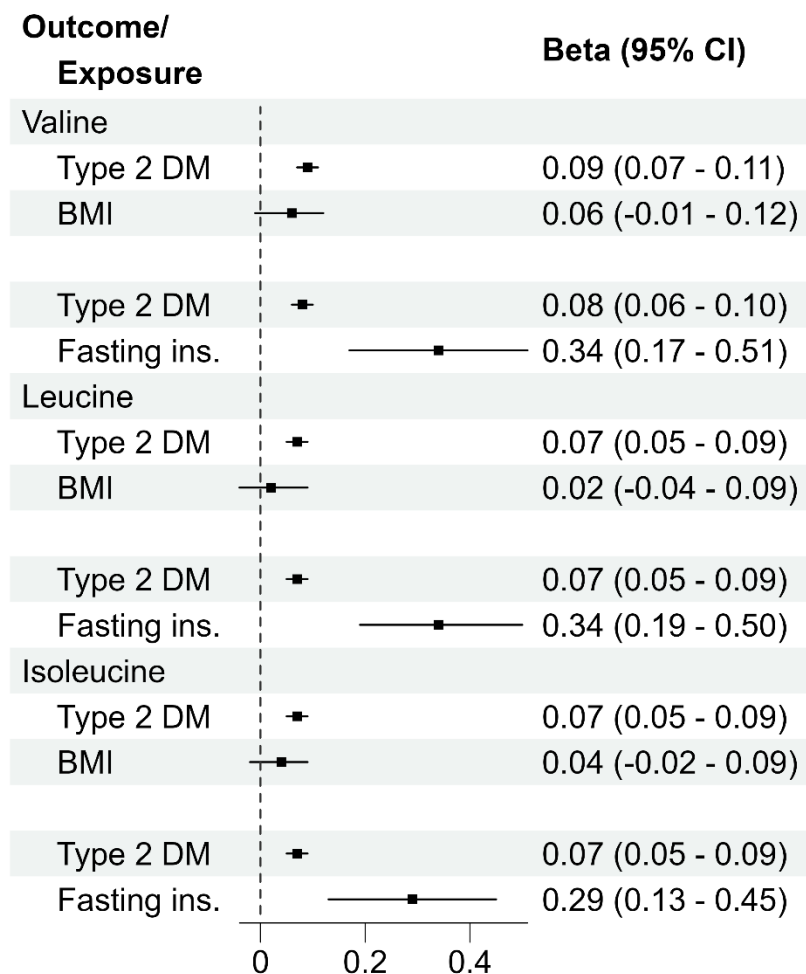
494 body mass index (BMI) (n=69 SNPs) and valine, leucine and isoleucine.

495



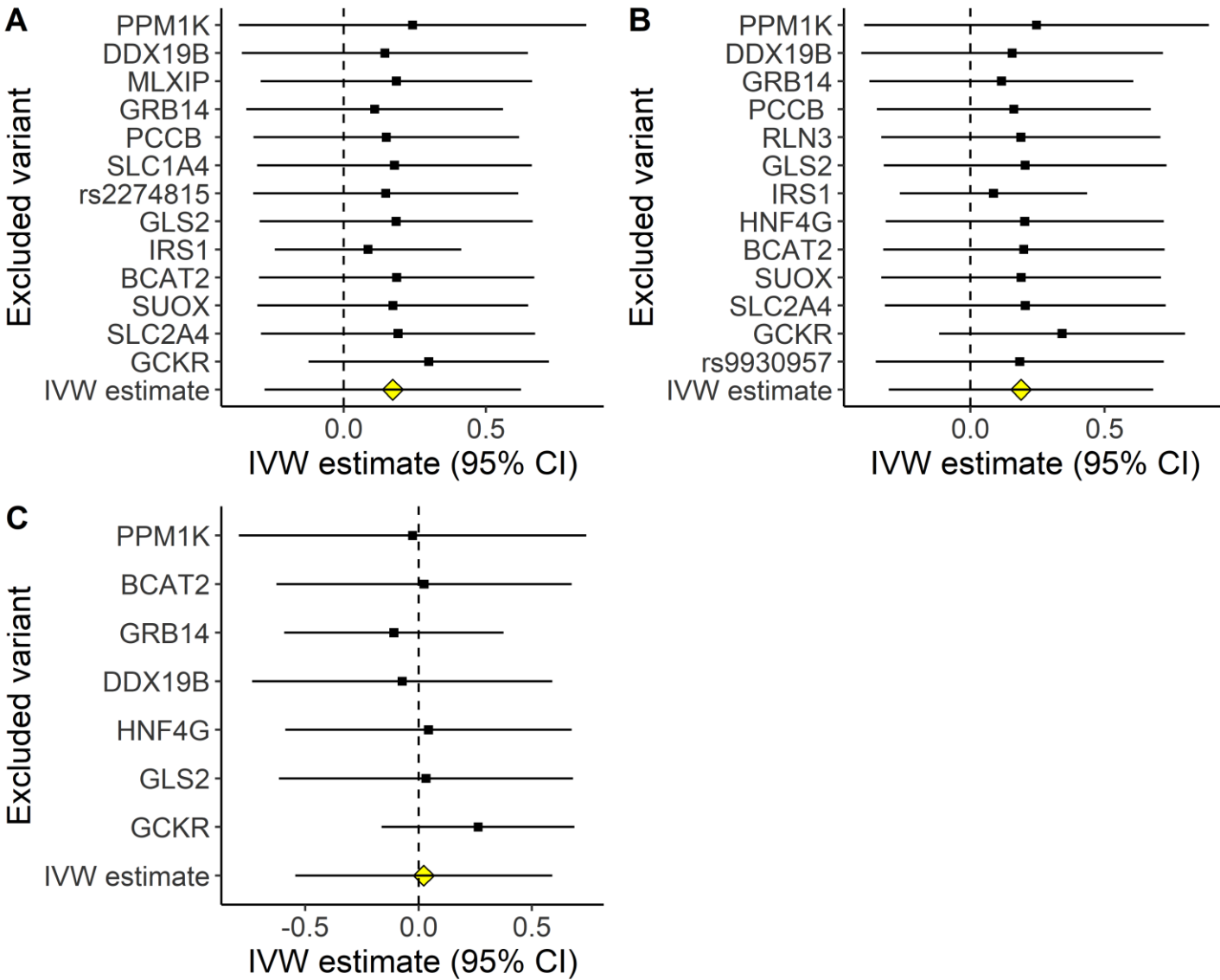
496

497 **Figure 5: Multivariable MR associations between type 2 diabetes and BCAAs.** Forest plots  
 498 summarizing the multivariable MR associations by the inverse variance weighted (IVW) method  
 499 between genetic instruments for type 2 diabetes and valine, leucine and isoleucine, adjusted for  
 500 either BMI or fasting insulin.  
 501



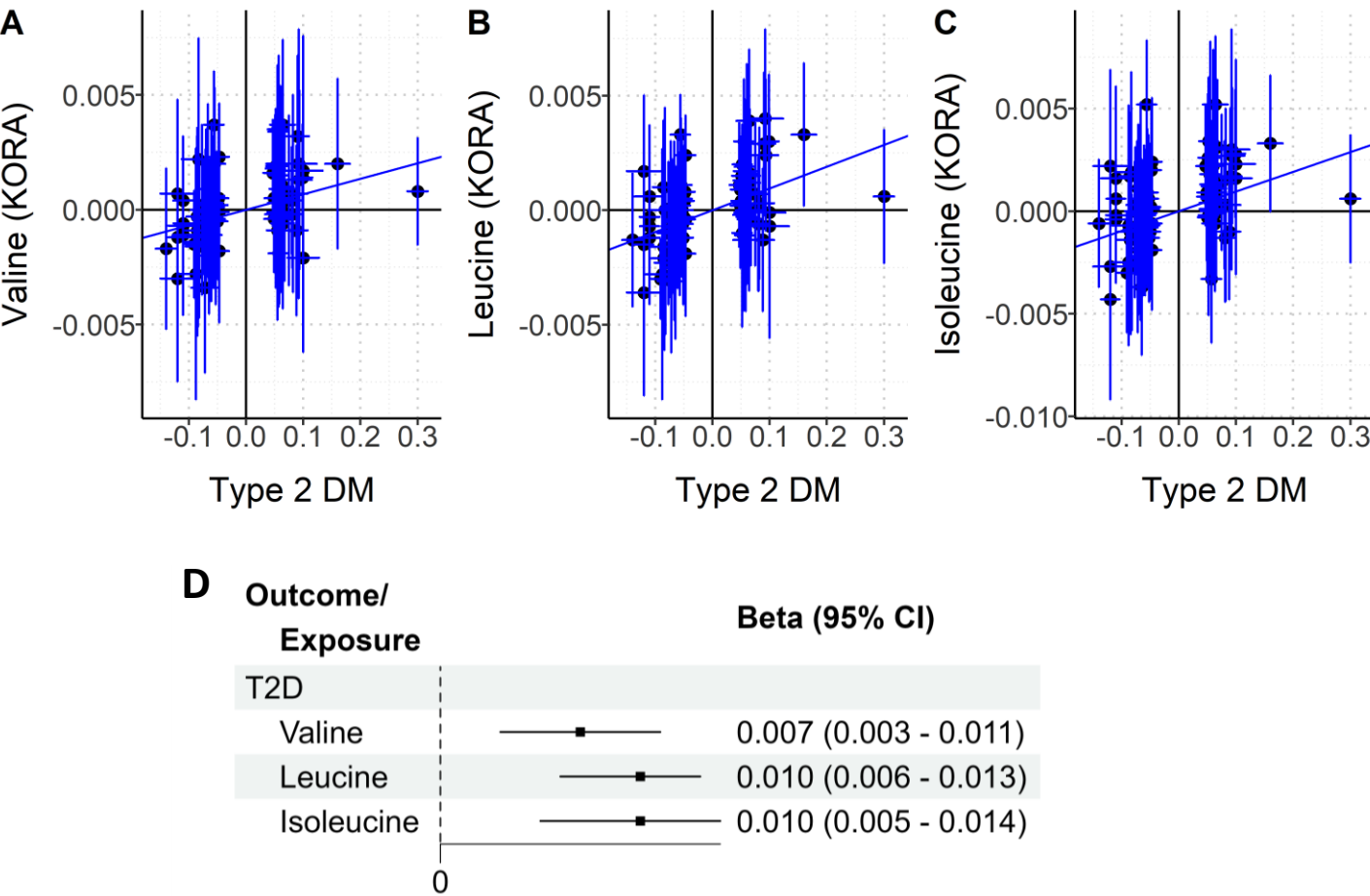
502  
 503  
 504

# Supplementary Figure 1



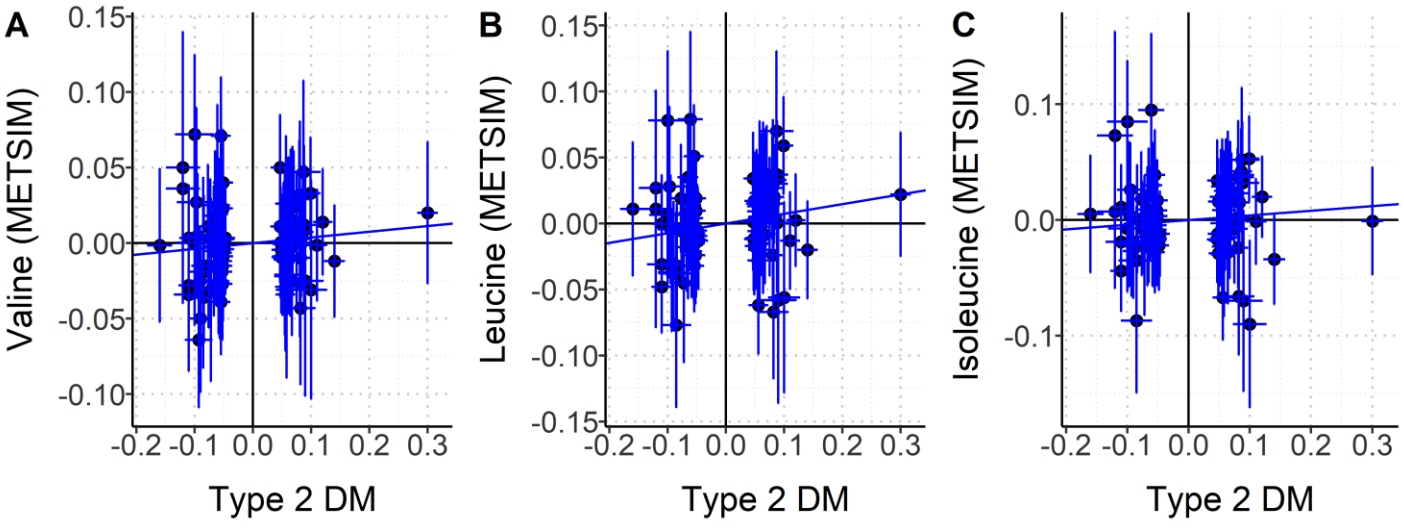
**Supplementary Figure 1.** Summary of a leave-one-out analysis for the associations between genetic instruments for (A) valine, (B) leucine and (C) isoleucine and type 2 diabetes. Each figures shows the IVW estimate (95% CI) after exclusion of the indicated SNP. Note that the name of the closest gene was substituted for the SNP name, when available.

# Supplementary Figure 2



**Supplementary Figure 2. Associations between SNP instruments for type 2 diabetes and BCAAs measured in the KORA population.** (A-C) Scatter plots showing of the associations between genetic instruments for type 2 diabetes (n=90 SNPs) and valine, leucine and isoleucine. The lines represent the association based on the IVW method. (D) Forest plot summarizing the associations by the inverse variance weighted (IVW) method between genetic instruments for type 2 diabetes and valine, leucine and isoleucine.

# Supplementary Figure 3

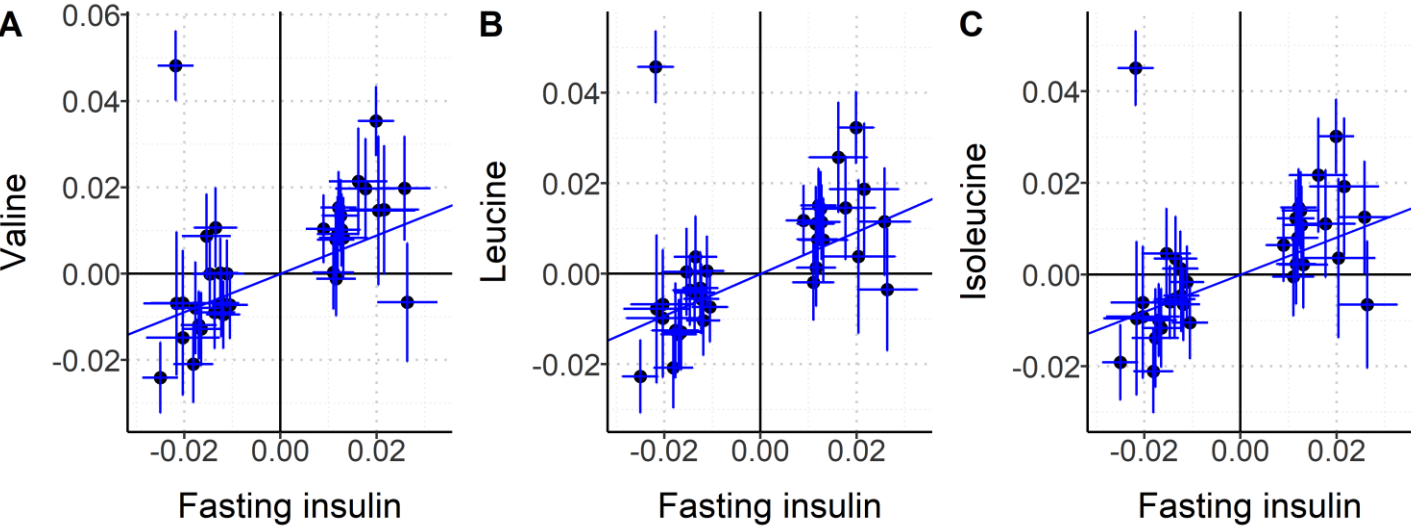


Outcome/ Exposure	Beta (95% CI)
T2D	
Valine	0.04 (-0.02 - 0.10)
Leucine	0.07 (0.01 - 0.14)
Isoleucine	0.04 (-0.03 - 0.11)

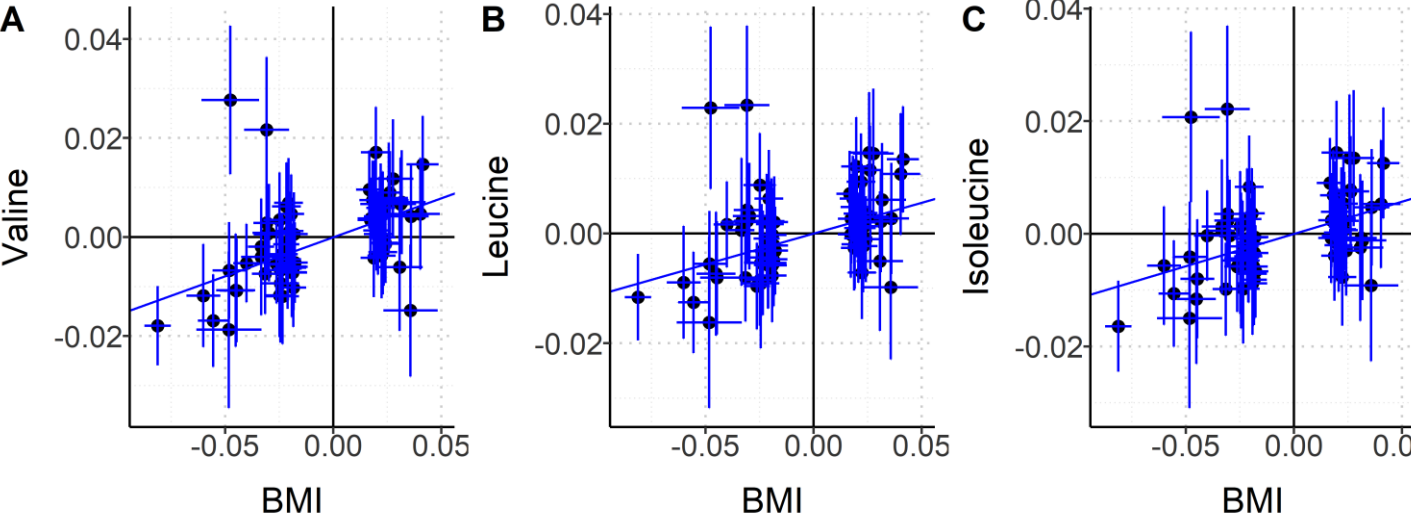
**Supplementary Figure 3. Associations between SNP instruments for type 2 diabetes and BCAAs measured in the METSIM population.** (A-C) Scatter plots showing of the associations between genetic instruments for type 2 diabetes (n=95 SNPs) and valine, leucine and isoleucine. The lines represent the association based on the IVW method. (D) Forest plot summarizing the associations by the inverse variance weighted (IVW) method between genetic instruments for type 2 diabetes and valine, leucine and isoleucine.

# Supplementary Figure 4

## Fasting insulin



## BMI associations



**Supplementary Figure 4. Associations between SNP instruments for fasting insulin or BMI, and the BCAAs measured in the UK biobank population.** Scatter plots showing of the associations between genetic instruments for fasting insulin (top panels) and BMI (bottom panels) and each BCAA. The lines represent the association based on the IVW method.