

1 **Metabolome-wide Mendelian randomization characterizes heterogeneous and**  
2 **shared causal effects of metabolites on human health**

3

4 Xianyong Yin<sup>1</sup>, Jack Li<sup>2</sup>, Debraj Bose<sup>2</sup>, Jeffrey Okamoto<sup>2</sup>, Annie Kwon<sup>2</sup>, Anne U. Jackson<sup>2</sup>,  
5 Lilian Fernandes Silva<sup>3</sup>, Anniina Oravilahti<sup>3</sup>, Heather M. Stringham<sup>2</sup>, Samuli Ripatti<sup>4,5,6</sup>,  
6 Mark Daly<sup>4,5,7</sup>, Aarno Palotie<sup>4,5,7</sup>, Laura J. Scott<sup>2</sup>, Charles F. Burant<sup>8</sup>, Eric B. Fauman<sup>9</sup>,  
7 Xiaoquan Wen<sup>2</sup>, Michael Boehnke<sup>2</sup>, Markku Laakso<sup>3</sup>, Jean Morrison<sup>2</sup>

8

9 <sup>1</sup>Department of Epidemiology, School of Public Health, Nanjing Medical University,  
10 Nanjing, Jiangsu 211166, China

11 <sup>2</sup>Department of Biostatistics and Center for Statistical Genetics, University of Michigan  
12 School of Public Health, Ann Arbor, MI 48109, USA

13 <sup>3</sup>Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and  
14 Kuopio University Hospital, Kuopio 70210, Finland

15 <sup>4</sup>Institute for Molecular Medicine Finland, FIMM, HiLIFE, University of Helsinki, Helsinki  
16 00290, Finland

17 <sup>5</sup>Department of Public Health, University of Helsinki, Helsinki 00014, Finland

18 <sup>6</sup>Broad Institute of MIT & Harvard, Cambridge, MA 02142, USA

19 <sup>7</sup>Analytic and Translational Genetics Unit, Department of Medicine, Department of  
20 Neurology, and Department of Psychiatry, Massachusetts General Hospital, Boston, MA  
21 02114, USA

22 <sup>8</sup>Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109, USA

23 <sup>9</sup>Internal Medicine Research Unit, Pfizer Worldwide Research, Development and Medical,  
24 Cambridge, MA 02139, USA

25

26 Correspondence: [xianyongyin@njmu.edu.cn](mailto:xianyongyin@njmu.edu.cn) (X.Y.), [boehnke@umich.edu](mailto:boehnke@umich.edu) (M.B.),  
27 [markku.laakso@uef.fi](mailto:markku.laakso@uef.fi) (M.L.), and [jymorr@umich.edu](mailto:jymorr@umich.edu) (J.M.)

28

29

30

31

32

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

### 33 **Summary**

34 Metabolites are small molecules that are useful for estimating disease risk and  
35 elucidating disease biology. Nevertheless, their causal effects on human diseases have not  
36 been evaluated comprehensively. We performed two-sample Mendelian randomization  
37 to systematically infer the causal effects of 1,099 plasma metabolites measured in 6,136  
38 Finnish men from the METSIM study on risk of 2,099 binary disease endpoints measured  
39 in 309,154 Finnish individuals from FinnGen. We identified evidence for 282 causal  
40 effects of 70 metabolites on 183 disease endpoints (FDR<1%). We found 25 metabolites  
41 with potential causal effects across multiple disease domains, including ascorbic acid 2-  
42 sulfate affecting 26 disease endpoints in 12 disease domains. Our study suggests that N-  
43 acetyl-2-aminooctanoate and glycocholate sulfate affect risk of atrial fibrillation  
44 through two distinct metabolic pathways and that N-methylpyrrolidone may mediate the  
45 causal effect of N6, N6-dimethyllysine on anxious personality disorder. This study  
46 highlights the broad causal impact of plasma metabolites and widespread metabolic  
47 connections across diseases.

48

## 49 **Introduction**

50 Metabolites are intermediate or end products of cellular metabolism with a wide range  
51 of functions<sup>1</sup>. Compared to gene transcripts and proteins, metabolites are more  
52 proximal to diseases, making them ideal biomarkers for estimating disease risk and  
53 understanding disease biology. Metabolite levels have shown associations with many  
54 human diseases, including type 2 diabetes and multiple cancers<sup>2,3</sup>. Some metabolites  
55 have demonstrated potential for predicting future disease<sup>4,5</sup>. However, the causal effects  
56 of metabolites on human diseases have not been evaluated comprehensively.

57 Metabolite levels reflect both environmental and genetic influences<sup>1</sup>. With the  
58 advent of high-throughput metabolic profiling technology, measuring levels of thousands  
59 of metabolites for participants in population studies has become possible. Recent  
60 genome-wide association studies (GWAS) that combine high-throughput metabolic  
61 profiling and genotyping/sequencing in large samples have identified thousands of  
62 genetic associations for thousands of metabolites and metabolic features<sup>6</sup>. These studies  
63 usually measure metabolite levels in blood, which are widely considered to reflect  
64 metabolite aggregate concentrations across tissues<sup>7</sup>. Recently, we profiled plasma levels  
65 for 1,391 metabolites using Metabolon non-targeted mass spectrometry technology in  
66 6,136 Finnish individuals of the Metabolic Syndrome in Men (METSIM) study<sup>8</sup>. GWAS  
67 identified 2,030 genetic associations for 803 of the 1,391 metabolites<sup>8</sup>. Integrating these  
68 metabolite GWAS with expression quantitative trait loci (eQTL) in 49 human tissues  
69 established associations of expression levels of 397 genes with levels of 521 plasma  
70 metabolites<sup>9</sup>. These GWAS deepen our understanding of genetic regulation of metabolic  
71 individuality, open an avenue to evaluate the causal effects of blood metabolites on  
72 human diseases using Mendelian randomization, and have the potential to provide  
73 actionable disease interventions.

74 Mendelian randomization is an instrumental variable (IV) method to interrogate  
75 causal effects of heritable risk factors on diseases of interest using genetic variants as  
76 IVs<sup>10</sup>. Mendelian randomization has identified modifiable risk factors for human diseases  
77 and recent methods development facilitates its broader application. For example,  
78 Mendelian randomization using the robust adjusted profile score (MR-RAPS) can account  
79 for bias of weak and outlier genetic IVs<sup>11</sup> and multivariable Mendelian randomization  
80 enables testing causal effects of multiple potentially related exposures on the same  
81 outcome<sup>12,13</sup>.

82 Mendelian randomization analysis has recently been applied to search for causal  
83 blood metabolites for a wide range of diseases and traits, including type 2 diabetes<sup>14</sup>,  
84 neuroticism<sup>15</sup>, Alzheimer’s disease<sup>16</sup>, and rheumatoid arthritis<sup>17</sup>. These studies  
85 demonstrate the utility of Mendelian randomization to identify potential causal  
86 metabolites and metabolic pathways for human diseases. However, the existing studies  
87 are restricted to one or a few disease outcomes and a relatively limited set of  
88 metabolites<sup>6,18</sup>.

89 Here, we comprehensively evaluated potential causal effects of 1,099 plasma  
90 metabolites on 2,099 binary disease endpoints (hereafter disease traits) using a  
91 Mendelian randomization analysis in GWAS of METSIM plasma metabolites<sup>8</sup> and FinnGen  
92 disease traits (release 7)<sup>19</sup>. We identified evidence for 282 causal effects of 70 plasma  
93 metabolites on 183 disease traits. Our study uncovered new potential causal effects of  
94 plasma metabolites for a broad spectrum of human diseases. We also identified some  
95 metabolites with broad causal effects across multiple disease types.

96

## 97 **Results**

98 **Interpretation of Mendelian randomization effect estimates.** Mendelian  
99 randomization tests whether genetic variants that affect the exposure (metabolite) have  
100 a proportional effect on the outcome (disease trait). With additional assumptions about  
101 the relationship between the genetic variants, metabolites, and disease traits<sup>20</sup>, the  
102 proportionality constant can be interpreted as a measure of the strength of the causal  
103 effect. In this paper, we focus primarily on significance and direction when interpreting  
104 estimated effects. Mendelian randomization can avoid bias due to environmental  
105 confounding and reverse causation which can plague observational associations<sup>20</sup>.  
106 However, causal interpretation of Mendelian randomization effects relies on additional  
107 assumptions, which may not hold in all cases. These effects must therefore be interpreted  
108 in the context of other sources of evidence (see Davies et al. 2017<sup>20</sup> for a full discussion  
109 of interpretation of Mendelian randomization estimates).

110

111 **Summary of Mendelian randomization results.** We previously conducted GWAS for  
112 1,099 named plasma metabolites with annotated chemical identities in up to 6,136  
113 Finnish men aged 45-74 at enrollment from the METSIM study<sup>8</sup>. These 1,099 metabolites  
114 included nine biochemical classes of small molecules related to the metabolisms of lipids

115 (n=548, 49.9%), amino acids (n=215, 19.6%), xenobiotics (n=163, 14.8%), peptides  
116 (n=42, 3.8%), nucleotides (n=42, 3.8%), cofactors and vitamins (n=38, 3.5%),  
117 carbohydrates (n=25, 2.3%), partially-characterized molecules (n=16, 1.5%), and energy  
118 (n=10, 0.9%) (**Supplementary Table 1**).

119 To identify causal plasma metabolites for human diseases, we carried out  
120 univariable Mendelian randomization analysis using MR-RAPS<sup>19</sup> to evaluate causal  
121 effects of the 1,099 metabolites on 2,099 binary disease traits from the FinnGen study  
122 (release 7; **Fig. 1a**). In GWAS, we inverse normalized the metabolite measurements<sup>8</sup> and  
123 measured disease trait associations by mixed-model logistic regression<sup>19</sup>. Our estimated  
124 causal effects can therefore be interpreted as the change in log odds of disease risk caused  
125 by an increase of one standard deviation of the normalized metabolite level. To identify  
126 independent IVs for the Mendelian randomization analysis, we performed linkage  
127 disequilibrium (LD) clumping in the GWAS summary statistics for each of the 1,099  
128 metabolites to ensure resulting variants achieve association  $P < 10^{-5}$  and each pair of  
129 variants within 1 megabase (Mb) distance satisfy  $LD\ r^2 < 0.01$ . For the 1,099 metabolites,  
130 we identified from 12 to 173 likely independent variants (mean=42.3; median=40.0) and  
131 used these as IVs (**Supplementary Fig. 1**).

132 We identified evidence for 282 causal effects of 70 plasma metabolites on 183  
133 disease traits at a false discovery rate (FDR) threshold  $< 1\%$  (**Fig. 2** and **Supplementary**  
134 **Table 2**), highlighting the broad relevance of plasma metabolite levels to human health.  
135 These 282 metabolite-disease trait pairs showed strong robustness to IV selection and  
136 choice of Mendelian randomization method (**Supplementary Fig. 2-5, Supplementary**  
137 **Notes**). The 70 causal metabolites comprised lipids (n=31, 44.3%), amino acids (n=29,  
138 41.4%), xenobiotics (n=4, 5.7%), cofactors and vitamins (n=2, 2.9%), and nucleotides,  
139 carbohydrate, peptide, and partially-characterized molecule (n=1, 1.4% for each).  
140 Compared to the 1,099 metabolites evaluated, the 70 metabolites showed significant  
141 enrichment in amino acids (odds ratio (OR)=3.20, Chi-square test  $P=4.0 \times 10^{-6}$ ) and  
142 depletion in xenobiotics (OR=0.33, Chi-square test  $P=0.041$ ), which may reflect the  
143 significantly larger numbers of IVs for amino acids than for xenobiotics (Student's t-test  
144  $P=1.2 \times 10^{-12}$ ). The 70 plasma metabolites conferred significant causal effects on 1 to 26  
145 disease traits (mean=4.0; median=1.0), with 32 (46%) showing significant causal effects  
146 on more than one disease trait (**Fig. 1b-1c**). The 183 disease traits covered a broad  
147 spectrum of diseases. The FinnGen consortium grouped these disease traits into 20

148 categories, including cancers (e.g. colon cancers), cardiometabolic (e.g. type 2 diabetes),  
149 infectious (e.g. tularaemia), neurological (e.g. Parkinson's disease), and mental and  
150 behavioral diseases (e.g. anxiety personality disorder) (**Supplementary Table 2**). Each  
151 of the 183 disease traits had 1 to 6 causal metabolites (mean=1.5; median=1.0); 53 (29%)  
152 had  $\geq 2$  causal metabolites (**Fig. 1d**).

153

154 **New potential causal metabolites for diseases.** Among the 282 causal effects, we  
155 reproduced several known relationships. For example, we identified a potential causal  
156 effect of low plasma lipid glycosyl-N-stearoyl-sphingosine levels on increasing risk of  
157 coronary artery disease ( $\beta=-0.11$ ,  $P=1.0\times 10^{-6}$ ), reinforcing the important role of  
158 sphingolipid metabolism in coronary artery disease<sup>21</sup>. Studies have reported high levels  
159 of valine, a branched-chain amino acid, associated with increased risk of type 2  
160 diabetes<sup>4,22</sup>. We validated, with nominal significance, the causal effect of plasma valine  
161 levels on risk of type 2 diabetes ( $\beta=0.041$ ,  $P=5.0\times 10^{-3}$ ). In addition, we found that elevated  
162 plasma N-acetylvaline levels decreased risk of type 2 diabetes ( $\beta=-0.085$ ,  $P=1.1\times 10^{-8}$ ). N-  
163 acetylvaline is a derivative of valine and belongs to a class of N-acyl-alpha amino acids.  
164 Multivariable Mendelian randomization including both valine and N-acetylvaline  
165 suggested that both metabolites have direct effects on type 2 diabetes (N-acetylvaline:  
166  $\beta=-0.096$ ,  $P=2.7\times 10^{-12}$ ; valine:  $\beta=0.087$ ,  $P=1.8\times 10^{-5}$ ), indicating a potentially important  
167 and complex role of valine metabolism in risk of type 2 diabetes. Interestingly, we found  
168 that high levels of two additional plasma N-acyl-alpha amino acids N-acetylglutamate  
169 ( $\beta=-0.11$ ,  $P=1.0\times 10^{-7}$ ) and N-acetylmethionine ( $\beta=-0.072$ ,  $P=5.5\times 10^{-7}$ ) potentially  
170 causally decreased risk of type 2 diabetes. The three N-acyl-alpha amino acids N-  
171 acetylvaline, N-acetylglutamate, and N-acetylmethionine show substantial phenotypic  
172 correlation and share many IVs (**Fig. 3, Supplementary Fig. 6**). For these three N-acyl-  
173 alpha amino acids, our GWAS previously identified genome-wide significant associations  
174 at the *ACY1* gene<sup>8</sup>, which encodes enzyme aminoacylase 1 that catalyzes the hydrolysis of  
175 acylated L-amino acids to L-amino acids. Mendelian randomization suggested that  
176 elevated plasma aminoacylase 1 levels<sup>23</sup> decreased levels of the three N-acyl-alpha amino  
177 acids ( $\beta<-1.20$ ,  $P<4.2\times 10^{-21}$ ) but increased risk of type 2 diabetes ( $\beta=0.16$ ,  $P=2.6\times 10^{-4}$ ),  
178 directionally consistent with the known function of aminoacylase 1 and a recently  
179 reported positive causal effect of aminoacylase 1 on type 2 diabetes<sup>24</sup>. These findings  
180 suggest a possible role of synthesis or degradation of N-acetylated proteins in type 2

181 diabetes. However, due to substantial sharing of IVs across the three N-acetyl amino acids,  
182 Mendelian randomization cannot identify whether this effect is due to one specific N-  
183 acetyl amino acid or multiple.

184 Our study also identified new potential causal metabolites for human diseases.  
185 Mendelian randomization recently suggested causal effects of plasma metabolites on risk  
186 of dementia<sup>16,25,26</sup>. We identified a significant potential protective effect of high plasma  
187 lipid 2-arachidonoyl-GPC (20:4) levels on risk of frontotemporal dementia ( $\beta=-0.89$ ,  
188  $P=1.2\times 10^{-6}$ ), a type of dementia characterized by progressive loss of neurons in the  
189 brain's frontal or temporal lobes. To the best of our knowledge, studies previously only  
190 reported 2-methoxyacetaminophen sulfate<sup>27</sup> with causal effect specifically on  
191 frontotemporal dementia. 2-arachidonoyl-GPC (20:4) is a lysophosphatidylcholine  
192 widely considered as a potent pro-inflammatory mediator<sup>28</sup>. Emerging evidence has  
193 demonstrated that neuroinflammation plays an important role in dementia<sup>29</sup>. Studies  
194 have identified a negative association of lysophosphatidylcholine with Alzheimer's  
195 disease<sup>30</sup>. Consistent with these results, we found a protective causal effect of increased  
196 2-arachidonoyl-GPC (20:4) levels on risk of frontotemporal dementia. We previously  
197 identified genome-wide associations for 2-arachidonoyl-GPC (20:4) around the  
198 *FADS1/FADS2*, two fatty acid desaturase genes<sup>8</sup>. Interestingly, we found that low  
199 expression of *FADS1/FADS2* in the whole blood but high expression in the brain  
200 significantly increased plasma 2-arachidonoyl-GPC (20:4) level<sup>9</sup>. *FADS1* variants could  
201 regulate erythrocyte arachidonic acid biosynthesis that subsequently induces  
202 inflammation in Alzheimer's disease<sup>31</sup>.

203 Chronic kidney disease affects >10% of the general population worldwide<sup>32</sup> and  
204 its risk factors are still poorly understood. We found evidence that elevated plasma  
205 xenobiotic sulfate levels increased risk of chronic kidney disease ( $\beta=0.080$ ,  $P=1.9\times 10^{-7}$ ).  
206 High sulfate levels have been previously found to be associated with disease progression  
207 and increased mortality in individuals with kidney disease<sup>33</sup>. Our previous GWAS  
208 identified a genome-wide significant association with plasma sulfate levels at the  
209 *SLC13A1* gene<sup>8</sup>, which encodes a sulfate transmembrane transporter and mediates the  
210 first step of sulfate absorption. *SLC13A1* is primarily expressed in the proximal renal  
211 tubules. We previously found that high expression of *SLC13A1* decreased plasma sulfate  
212 abundance<sup>9</sup>. These results together suggest that *SLC13A1* could serve as a potential drug  
213 target for chronic kidney disease through regulation of plasma sulfate levels.

214

215 **Causal metabolites shared across diseases.** We identified evidence for 32 metabolites  
216 with causal effects on more than one disease trait (**Fig. 1b** and **Fig. 2**; see **Summary of**  
217 **Mendelian randomization results**). Of these 32 metabolites, 25 (78%) showed  
218 significant causal effects on  $\geq 2$  distinct disease categories (**Fig. 1c**, **Supplementary**  
219 **Table 2**). The sharing of causal metabolites between diseases may partially explain  
220 observed phenotypic correlations and disease comorbidities. For example, we identified  
221 causal effects of plasma amino acid N-acetylvaline levels on optic atrophy ( $\beta=0.53$ ,  
222  $P=4.7\times 10^{-7}$ ) and myasthenia gravis ( $\beta=0.53$ ,  $P=7.9\times 10^{-8}$ ), diseases with substantial  
223 comorbidity<sup>34</sup>. These results suggested that valine metabolism might play a role in both  
224 cell cycle of retinal ganglion cell axons and communication between nerves and muscle.  
225 We found causal effects of plasma amino acid N-acetyl-aspartyl-glutamate (NAAG) levels  
226 on increased risk of both Parkinson's disease ( $\beta=0.11$ ,  $P=3.2\times 10^{-7}$ ) and autoimmune  
227 hypothyroidism ( $\beta=0.039$ ,  $P=3.9\times 10^{-9}$ ), which also have substantial comorbidity<sup>35</sup>. To the  
228 best of our knowledge, this is the first reported evidence of these four causal effects.

229 The metabolite affecting the largest number of disease traits was ascorbic acid 2-  
230 sulfate, with evidence of causal effects on 26 disease traits of 12 categories, including  
231 cardiomyopathy (disease of the circulatory system), arthropathy (disease of the  
232 musculoskeletal system and connective tissue), and acne (disease of the skin and  
233 subcutaneous tissue) (**Supplementary Table 2**). We found that elevated levels of  
234 ascorbic acid 2-sulfate may decrease risk of 12 disease traits including colon  
235 adenocarcinoma ( $\beta=-0.13$ ;  $P=9.3\times 10^{-8}$ ) and endometriosis of the fallopian tube ( $\beta=-0.48$ ;  
236  $P=1.6\times 10^{-7}$ ) but increase risk of 14 others including conjunctiva cancer ( $\beta=0.36$ ;  
237  $P=2.8\times 10^{-14}$ ) and arthropathy ( $\beta=0.028$ ;  $P=1.1\times 10^{-7}$ ).

238 Notably, the suggested causal effects of plasma ascorbic acid 2-sulfate showed  
239 heterogeneity across disease traits even in the same category. For example, we found  
240 elevated ascorbic acid 2-sulfate levels are protective for acne ( $\beta=-0.18$ ;  $P=3.9\times 10^{-10}$ ) and  
241 lichen sclerosus ( $\beta=-0.15$ ;  $P=7.1\times 10^{-7}$ ) but increase risk of dyshidrosis, a kind of eczema  
242 ( $\beta=0.42$ ;  $P=4.2\times 10^{-10}$ ). These three conditions all affect skin but usually in different  
243 anatomical locations: the face, upper part of the chest, and back; the genital area; and the  
244 palms and fingers, respectively. Ascorbic acid 2-sulfate arises from the action of a liver-  
245 derived sulfotransferase on vitamin C, so it is possible that plasma levels of ascorbic acid  
246 2-sulfate are a proxy for action of liver-derived sulfotransferases or for vitamin C levels,



247 or a combination of these. Vitamin C is an essential nutrient for humans, acting as an  
248 antioxidant by protecting the body against oxidative stress, as a cofactor in enzymatic  
249 reactions including collagen synthesis, and as a structure component for blood vessels,  
250 cartilage, and muscle<sup>36</sup>. Vitamin C supplementation has been broadly recommended to  
251 help protect cells against the effects of free radicals, and has generally been found to be  
252 safe. Further investigation is needed to understand whether the effects we identified are  
253 effects of vitamin C itself or other biological processes.

254

255 **Independent causal metabolic pathways on the same disease.** We computed both  
256 phenotypic correlation and correlation of IV effects ( $r_{IV}$ ) for each pair of the 70 significant  
257 metabolites, showing their pervasive connections (**Fig. 3, Supplementary Fig. 7-8; see**  
258 **Methods**). We found strong correlations between some pairs of potential causal  
259 metabolites for the same disease traits (absolute  $r_{IV}$  median=0.84, mean=0.64,  
260 range=0.00033-0.99; **Supplementary Fig. 9**).

261 Causal effects of two metabolites with highly correlated IVs on the same disease  
262 trait in univariable Mendelian randomization could result from multiple scenarios. For  
263 example, both metabolites may causally affect the disease trait independently, or only  
264 one metabolite could affect the disease trait, with the result for the other being due to  
265 mediation or horizontal pleiotropy. We employed multivariable Mendelian  
266 randomization<sup>13</sup> to distinguish these possibilities.

267 For atrial fibrillation, we identified a risk effect of plasma lipid N-acetyl-2-amino-  
268 octanoate ( $\beta=0.068$ ;  $P=2.3\times 10^{-7}$ ) and protective effects of plasma amino acid N-delta-  
269 acetylornithine ( $\beta=-0.047$ ;  $P=5.1\times 10^{-7}$ ) and lipid glycocholate sulfate ( $\beta=-0.061$ ;  
270  $P=2.9\times 10^{-8}$ ). N-acetyl-2-amino-octanoate and N-delta-acetylornithine have highly  
271 correlated IVs ( $r_{IV}=0.74$ ) but neither has correlated IVs with glycocholate sulfate  
272 ( $|r_{IV}|<0.08$ ). Multivariable Mendelian randomization analysis identified distinct causal  
273 effects on atrial fibrillation of lipids N-acetyl-2-amino-octanoate ( $\beta=0.054$ ;  $P=7.2\times 10^{-3}$ )  
274 and glycocholate sulfate ( $\beta=-0.058$ ;  $P=2.6\times 10^{-7}$ ), but no causal effect of N-delta-  
275 acetylornithine, conditional on the other two metabolites ( $\beta=-0.020$ ;  $P=0.17$ ). In the  
276 METSIM study, we identified 816 individuals with atrial fibrillation (see **Methods**).  
277 Logistic regression identified a significant association between plasma N-acetyl-2-amino-  
278 octanoate level and risk of atrial fibrillation ( $\beta=0.080$ ;  $P=0.045$ ), directionally consistent  
279 with the causal effect estimated in Mendelian randomization. We observed no significant

280 associations with N-delta-acetylornithine ( $\beta=0.057$ ;  $P=0.148$ ) or glycocholate sulfate  
281 levels ( $\beta=0.072$ ;  $P=0.064$ ), however observational associations may be biased by  
282 unmeasured confounding variables.

283 For anxious personality disorder, we identified risk effects of plasma xenobiotic  
284 N-methylpiperolate ( $\beta=0.28$ ;  $P=2.8\times 10^{-7}$ ) and amino acid N6, N6-dimethyllysine ( $\beta=0.24$ ;  
285  $P=8.6\times 10^{-8}$ ) and a protective effect of plasma lipid androsterone sulfate ( $\beta=-0.27$ ;  
286  $P=1.5\times 10^{-7}$ ). N6,N6-dimethyllysine and N-methylpiperolate have high IV correlation  
287 ( $r_{IV}=0.98$ ) and share 42.4% of their IVs at a threshold of metabolite association  $P\leq 1\times 10^{-5}$ ,  
288 but neither has correlated IVs with androsterone sulfate ( $|r_{IV}|<0.03$ ). Because of the  
289 high IV correlation between N6, N6-dimethyllysine and N-methylpiperolate, there is not  
290 enough independent genetic signal to tease apart their causal effects on anxious  
291 personality disorder using multivariable Mendelian randomization. We performed two  
292 multivariable Mendelian randomization analyses including androsterone sulfate and  
293 either N-methylpiperolate or N6, N6-dimethyllysine. In both cases, the data to be  
294 consistent with direct effects of both included metabolites (N-methylpiperolate ( $\beta=0.29$ ;  
295  $P=6.2\times 10^{-8}$ ) and androsterone sulfate ( $\beta=-0.27$ ;  $P=7.6\times 10^{-8}$ ) or at N6, N6-dimethyllysine  
296 ( $\beta=0.24$ ;  $P=5.0\times 10^{-7}$ ) and androsterone sulfate ( $\beta=-0.27$ ;  $P=2.5\times 10^{-7}$ )). To further  
297 understand this relationship, we carried out a GWAS on the metabolite ratio of N6,N6-  
298 dimethyllysine and N-methylpiperolate, identifying six independent association signals  
299 in the *AKR1C1/AKR1C2/AKR1C3/AKR1C4/AKR1C8*, *NAT8*, *PYROXD2*, *SLC6A20*, and  
300 *SLC7A9* regions ( $P<5.0\times 10^{-8}$ ) (**Supplementary Table 3, Supplementary Fig. 10**).  
301 Mendelian randomization identified evidence for a causal effect of increased N6,N6-  
302 dimethyllysine:N-methylpiperolate ratio on risk of anxious personality disorder ( $\beta=-0.34$ ;  
303  $P=0.047$ ; **Supplementary Fig. 11**; see **Methods**). The pattern we observe in which N6,  
304 N6-dimethyllysine and N-methylpiperolate both increase risk of anxious personality  
305 disorder, but an increase in their ratio confers a protective effect is consistent with a  
306 hypothesis that N-methylpiperolate acts as a mediator in the potential causal pathway of  
307 N6, N6-dimethyllysine on anxious personality disorder (**Fig. 4**). This is consistent with  
308 previous reports that piperolate is an intermediate product of lysine metabolism<sup>37</sup>.

309

## 310 Discussion

311 In this study, we systematically screened for potential causal effects of 1,099 plasma  
312 metabolites on 2,099 disease endpoints using two-sample univariable and multivariable

313 Mendelian randomization analysis. We identified evidence for 282 causal effects of 70  
314 plasma metabolites on 183 disease endpoints. We characterized the sharing of  
315 metabolite causal effects across 53 human diseases and showed the heterogeneity of  
316 causal metabolic pathways in disease pathophysiology. This study uncovers modifiable  
317 risk metabolites for disease intervention and underscores a pervasive potential causal  
318 role of plasma metabolites in human health.

319 We identified evidence for causal effects of 70 plasma metabolites on 183 human  
320 diseases. The relationships of many plasma metabolites with diseases have not been  
321 studied previously. These findings have several implications. First, they provide potential  
322 targets for disease intervention. Many plasma metabolites levels can be modified by diet  
323 and lifestyle changes. For example, we identified that high plasma sulfate levels increased  
324 risk of chronic kidney disease. A wide range of food and beverages have been suggested  
325 as sources of dietary sulfate. We can, in principle, reduce plasma sulfate levels by  
326 reducing the consumption of these food and beverages.

327 Second, these findings help elucidate disease biology and prioritize therapeutic  
328 targets for human diseases. For example, the risk of high plasma sulfate on chronic kidney  
329 disease suggested *SLC13A1* as a potential drug target for chronic kidney disease. The  
330 protective effect of high 2-arachidonoyl-GPC (20:4) level on frontotemporal dementia  
331 bolsters the hypothesis that neuroinflammation contributes to the pathophysiology of  
332 dementia<sup>29,31</sup>. We characterize the pervasive sharing of potential causal metabolites and  
333 their heterogeneity effects across human diseases. The sharing may help explain some  
334 disease comorbidity and reveal previously-unappreciated connections between diseases.  
335 For example, we identified evidence for 126 heterogeneous causal effects of 15 N-acyl-  
336 alpha amino acids on 67 disease traits of 14 categories, highlighting a broad impact of  
337 synthesis or degradation of N-acetylated proteins on human health.

338 Our study showed that metabolites with significant univariable causal effects on  
339 the same disease traits might act in disease pathogenesis through separate metabolic  
340 pathways or through a metabolic cascade. We identified two independent metabolic  
341 pathways among three tested metabolites for atrial fibrillation and for anxious  
342 personality disorder, highlighting the heterogeneity of potential causal metabolic  
343 pathways in human diseases. We suggested that a causal effect of N-delta-acetylor  
344 on atrial fibrillation might be induced by IVs shared with N-acetyl-2-amino-octanoate. In  
345 contrast, we suggested that N-methylpipecolate might act as a downstream mediator in

346 the causal pathway of N6, N6-dimethyllysine on anxious personality disorder, which  
347 could partially explain the strong IV correlation between N-methylpipecolate and N6, N6-  
348 dimethyllysine. Previous survival analyses detected significant positive association of  
349 glycocholate sulfate levels with atrial fibrillation incidence<sup>38</sup>, while our analysis  
350 identified a negative association of plasma glycocholate sulfate with atrial fibrillation.  
351 The effect of plasma glycocholate sulfate on atrial fibrillation warrants further  
352 investigation.

353 In this study, plasma metabolite levels may act as proxies for activity of specific  
354 biological pathways, or levels of metabolites in other tissues. For ease of exposition, we  
355 referred to causal effects of metabolites on disease traits throughout this report. However,  
356 this may not mean that intervening directly on plasma metabolite levels will impact risk  
357 of disease trait. The biochemical pathway regulating the metabolite may be the true  
358 causal culprit. For example, we identified a negative association of plasma 2-  
359 arachidonoyl-GPC (20:4) level with the risk of frontotemporal dementia, which might  
360 suggest a role of 2-arachidonoyl-GPC (20:4)-mediated neuroinflammation in the brain. In  
361 addition, we applied multivariable Mendelian randomization to tease apart independent  
362 potential causal effects of metabolites on the same disease. However, multivariable  
363 Mendelian randomization can only distinguish effects of metabolites that have a sufficient  
364 number of distinct IVs. Finally, associations between IVs and metabolites were estimated  
365 in an all-male cohort<sup>8</sup>, while IV-disease associations were estimated in a mixed cohort.  
366 Our causal effect estimates rely on the assumption that genetic regulation of metabolites  
367 and causal effects do not differ between the sexes. If these assumptions are violated, our  
368 estimates will be inaccurate or may not generalize to a mixed sex population. This issue  
369 is most likely to affect sexually differentiated metabolites such as androsterone sulfate.

370 In conclusion, we systematically evaluated the causal effects of 1,099 plasma  
371 metabolites on the risk of 2,099 disease endpoints. We identified evidence for 282 causal  
372 effects of 70 plasma metabolites on 183 disease traits. Our study newly uncovered  
373 potential causal effects of plasma metabolites on a broad spectrum of human diseases.  
374 These findings highlight heterogeneous and shared causal effects of plasma metabolites  
375 on human diseases.

376  
377  
378

379 **Methods**

380 **Metabolic Syndrome In Men (METSIM) metabolomics study.** METSIM is a single-site  
381 cohort study designed to investigate risk factors for type 2 diabetes and cardiovascular  
382 diseases<sup>39</sup>. It includes 10,197 Finnish men from Kuopio aged 45 to 74 years at baseline.  
383 We performed non-targeted metabolomics profiling in 6,136 randomly-selected non-  
384 diabetic participants using the Metabolon DiscoveryHD4 mass spectrometry platform  
385 (Durham, North Carolina, USA) on EDTA-plasma samples obtained after ≥10-hour  
386 overnight fast during baseline visits from 2005 to 2010<sup>8</sup>. We completed single-variant  
387 GWAS for 1,391 metabolites, which identified 2,030 independent metabolite  
388 associations<sup>8</sup>. For this study, we used GWAS summary statistics at 16.2M genotyped or  
389 imputed genetic variants for the 1,099 named metabolites with annotated biochemical  
390 identities<sup>8</sup>. All METSIM participants provided written informed consent. The Ethics  
391 Committee at the University of Eastern Finland and the Institutional Review Board at the  
392 University of Michigan approved the METSIM metabolomics study.

393

394 **FinnGen study.** FinnGen is designed to collect and analyze genome and healthcare data  
395 to identify new diagnostic and therapeutic targets for human diseases<sup>19</sup>. FinnGen  
396 obtained participant informed consent for biobank research based on the Finnish  
397 Biobank Act. Research cohorts collected prior to the Finnish Biobank Act coming into  
398 effect (September 2013) and the start of FinnGen (August 2017) obtained study-specific  
399 consents and later transferred the consents to the Finnish biobank after the National  
400 Supervisory Authority for Welfare and Health (Fimea) approved the recruitment  
401 protocols.

402 FinnGen identified 3,095 disease endpoints in release 7 using healthcare data  
403 from Finnish national registries: Drug Purchase and Drug Reimbursement and Digital and  
404 Population Data Services Agency; Digital and Population Data Services Agency; Statistics  
405 Finland; Register of Primary Health Care Visits (AVOHILMO); Care Register for Health  
406 Care (HILMO); and Finnish Cancer Registry. These registries recorded disease-relevant  
407 codes of the International Classification of Diseases (ICD) revisions 8, 9, and 10, cancer-  
408 specific ICD-O-3, Nordic Medico-Statistical Committee (NOMESCO) procedure, Finnish-  
409 specific Social Insurance Institute (KELA) drug reimbursement, and Anatomical  
410 Therapeutic Chemical (ATC)<sup>8</sup>. Each FinnGen participant was genotyped with an Illumina  
411 or Affymetrix array. Genotype imputation followed using the Finnish-specific Sequencing

412 Initiative Suomi (SISu) v3 reference panel<sup>40</sup>. FinnGen carried out single-variant GWAS for  
413 each disease endpoint using mixed model logistic regression in SAIGE<sup>41</sup>. For this study,  
414 we used GWAS summary statistics at 16.7M genotyped or imputed genetic variants for  
415 all 3,095 disease traits in up to 309,154 individuals from FinnGen release 7. After we  
416 finished the Mendelian randomization analysis, FinnGen made the release 8 publicly  
417 available, which includes GWAS summary statistics for 2,202 disease traits. In  
418 comparison to FinnGen release 7, release 8 reduced the number of disease traits  
419 primarily by dropping redundant disease traits. To improve efficiency and reduce  
420 redundancy, we restricted our Mendelian randomization analysis results to 2,099 of the  
421 3,095 disease traits that are included in FinnGen release 8.

422

423 **Selection of IVs.** We identified 16.2M genetic variants shared between GWAS summary  
424 files across all the 1,099 metabolites in METSIM and the 2,099 disease traits in FinnGen  
425 release 7. To identify independent genetic variants as IVs for Mendelian randomization,  
426 we performed LD clumping in the GWAS results for each of the 1,099 metabolites in Plink  
427 to ensure resulting variants achieved association  $P < 10^{-5}$  and each pair of variants within  
428 1 Mb distance has LD  $r^2 < 0.01$ <sup>42</sup>. For LD calculation, we used genotypes in 8,433 METSIM  
429 individuals without close relatives defined as pairwise kinship coefficients  $< 0.125$ .

430

431 **Primary univariable Mendelian randomization.** To identify causal metabolites for  
432 human diseases, we performed two-sample univariable Mendelian randomization to test  
433 the causal effect of each of the 1,099 plasma metabolites on each of the 2,099 disease  
434 traits using MR-robust adjusted profile scoring (MR-RAPS)<sup>11</sup>. MR-RAPS allows for  
435 horizontal pleiotropy and enables inclusion of IVs with weak effects by accounting for the  
436 precision of IV-exposure and IV-outcome associations<sup>11</sup>. We used over dispersion and  
437 Tukey robust loss function parameters in MR-RAPS. We conducted the MR-RAPS analysis  
438 using the `mr.raps` R package. To identify significant causal effects, we applied an FDR  $< 1\%$   
439 to account for multiple testing.

440

441 **Evaluation of causal effects of blood aminoacylase 1 levels on plasma levels of three**  
442 **N-acyl-alpha amino acids and risk of type 2 diabetes.** To test causal effects of protein  
443 aminoacylase 1 on plasma levels of three N-acyl-alpha amino acids: N-acetylvaline, N-  
444 acetylglutamate, and N-acetylmethionine and risk of type 2 diabetes, we performed two-

445 sample univariable Mendelian randomization. deCODE measured plasma aminoacylase 1  
446 level using SomaScan version 4 in 35,559 Icelanders followed by protein quantitative  
447 trait loci (pQTL) analysis, which identified three independent cis-pQTLs for aminoacylase  
448 <sup>123</sup>. Among the three cis-pQTLs, the top pQTL site rs121912698 was available in both  
449 METSIM and FinnGen. We used this variant as single IV and performed a Wald ratio test  
450 to evaluate causal effects of protein aminoacylase 1 on plasma levels of the three N-acyl-  
451 alpha amino acids and risk of type 2 diabetes in the twoSampleMR R package.

452

453 **Estimation of IV correlation between metabolites.** To estimate the degree to which  
454 each pair of metabolites share genetic IVs, we computed the proportion of overlapping  
455 IVs and the IV correlation. For each metabolite pair, we took the union of IVs for both  
456 metabolites. We then performed LD clumping using LD  $r^2 < 0.01$  in 1 Mb distance in Plink<sup>42</sup>  
457 to remove correlated IVs. Finally, we extracted association statistics for the resulting set  
458 of IVs for both metabolites. For LD calculation, we used genotypes in 8,433 METSIM  
459 individuals with pairwise kinship coefficients  $< 0.125$ . We calculated the proportion of IVs  
460 shared as the proportion of the LD clumped union set of IVs with association  $P \leq 10^{-5}$  for  
461 both metabolites. We calculated the IV correlation,  $r_{IV}$ , as the correlation of association  
462 statistics of the LD clumped union set of IVs with the two metabolites.

463

464 **Multivariable Mendelian randomization.** To detect independent causal effects among  
465 metabolites that conferred significant univariable causal effects on the same disease trait,  
466 we performed multivariable Mendelian randomization in Genome-wide mR Analysis  
467 under Pervasive PLEiotropy (GRAPPLE)<sup>13</sup>. We merged the IVs that were used in  
468 univariable Mendelian randomization across all the targeted metabolites and performed  
469 LD clumping as in **Selection of IVs** to ensure that all IVs were nearly independent. We  
470 applied default parameters in GRAPPLE and used nominal  $P < 0.05$  as the significance  
471 threshold.

472

473 **Associations of N-acetyl-2-aminooctanoate, N-delta-acetylornithine, and**  
474 **glycocholenate sulfate with atrial fibrillation in METSIM.** Among the 6,102 METSIM  
475 participants with measured plasma N-acetyl-2-aminooctanoate, N-delta-acetylornithine,  
476 and glycocholenate sulfate levels at baseline, we identified 816 with atrial fibrillation in  
477 METSIM as of June 2022. To test for associations between plasma metabolite levels and

478 presence of atrial fibrillation, we used logistic regression with covariates baseline study  
479 age, body mass index (BMI), binary cigarette smoking status (ever smoker versus never  
480 smoker), alcohol drinking amount, baseline systolic and diastolic blood pressure, and  
481 lipid and hypertension medication use.

482

483 **GWAS for metabolite ratio of N6,N6-dimethyllysine and N-methylpipecolate and**  
484 **causal effect of the ratio on anxious personality disorder.** In the 6,136 METSIM  
485 participants<sup>8</sup>, we computed the ratio of N6,N6-dimethyllysine to N-methylpipecolate by  
486 dividing the level of N6,N6-dimethyllysine by the level of N-methylpipecolate. We  
487 regressed out covariates study age, Metabolon batches, and lipid lowering medication  
488 status, and inverse normalized the residuals. We performed single-variant GWAS for the  
489 resulting residuals in Regenie v3.2.2<sup>43</sup>. For the chromosomes on which we identified  
490 genome-wide significant associations ( $P < 5.0 \times 10^{-8}$ ), we performed recursively a stepwise  
491 conditional test to identify near-independent association signals until no variant attained  
492  $P < 5.0 \times 10^{-8}$ . To test causal effect of the metabolite ratio on risk of anxious personality  
493 disorder, we performed univariable Mendelian randomization test using MR-RAPS<sup>11</sup>. We  
494 used the near-independent association signals for the metabolite ratio that are also  
495 available in the GWAS for anxious personality disorder as IVs. We conducted the MR-  
496 RAPS analysis with over dispersion and Tukey robust loss function parameters using the  
497 mr.raps R package.

498



## 499 **Acknowledgements**

500 We thank all the participants and investigators in the METSIM and FinnGen studies. This  
501 work was supported by the National Institutes of Health (NIH) under awards U01  
502 DK062370 (M.B.), R35 GM138121 (X.Q.W.), R01 DK119380 (X.Q.W.), the American  
503 Diabetes Association Postdoctoral Fellowship (1-19-PDF-061, X.Y.), the University of  
504 Michigan Precision Health Scholarship (X.Y.), the Academy of Finland under grant no.  
505 321428 (M.L.), the Sigrid Juselius Foundation (M.L.), the Academy of Finland Center of  
506 Excellence in Complex Disease Genetics under grant no. 312062 and 336820 (S.R.), grant  
507 no. 312074 and 336824 (A.P.), the Finnish Foundation for Cardiovascular Research (S.R.),  
508 University of Helsinki HiLIFE Fellow and Grand Challenge grants, and Horizon 2020  
509 Research and Innovation Programme (grant no. 101016775 “INTERVENE”) (S.R.).

510 The Coordinating Ethics Committee of the Hospital District of Helsinki and  
511 Uusimaa (HUS) approved the FinnGen study protocol Nr HUS/990/2017. The FinnGen  
512 study is approved by Finnish Institute for Health and Welfare (permit numbers:  
513 THL/2031/6.02.00/2017, THL/1101/5.05.00/2017, THL/341/6.02.00/2018,  
514 THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019,  
515 THL/1524/5.05.00/2020, and THL/2364/14.02/2020), Digital and population data  
516 service agency (permit numbers: VRK43431/2017-3, VRK/6909/2018-3,  
517 VRK/4415/2019-3), Social Insurance Institution (permit numbers: KELA 58/522/2017,  
518 KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, KELA 138/522/2019,  
519 KELA 2/522/2020, KELA 16/522/2020), and Statistics Finland (permit numbers: TK-53-  
520 1041-17 and TK-53-90-20).

521

## 522 **Declaration of interests**

523 E.B.F. is an employee and stockholder of Pfizer. The remaining authors declare no  
524 competing interests.

525

## 526 **Web resources**

527 METSIM metabolomics PheWeb: <https://pheweb.org/metsim-metab>

528 FinnGen: <https://www.finnngen.fi>

529 FinnGen documentation: <https://finngen.gitbook.io/documentation>

530

## 531 **Data and code availability**

532 FinnGen genome-wide summary statistics are available at <https://r7.finngen.fi>.  
533 Full summary statistics from the genome-wide association studies of the 1,099 plasma  
534 metabolites are available at <https://pheweb.org/metsim-metab/>. TwoSampleMR is  
535 available at <https://github.com/MRCIEU/TwoSampleMR>. MR-RAPS is available  
536 <https://github.com/qingyuanzhao/mr.raps>. GRAPPLE is available at  
537 <https://github.com/jingshuw/GRAPPLE>.

538

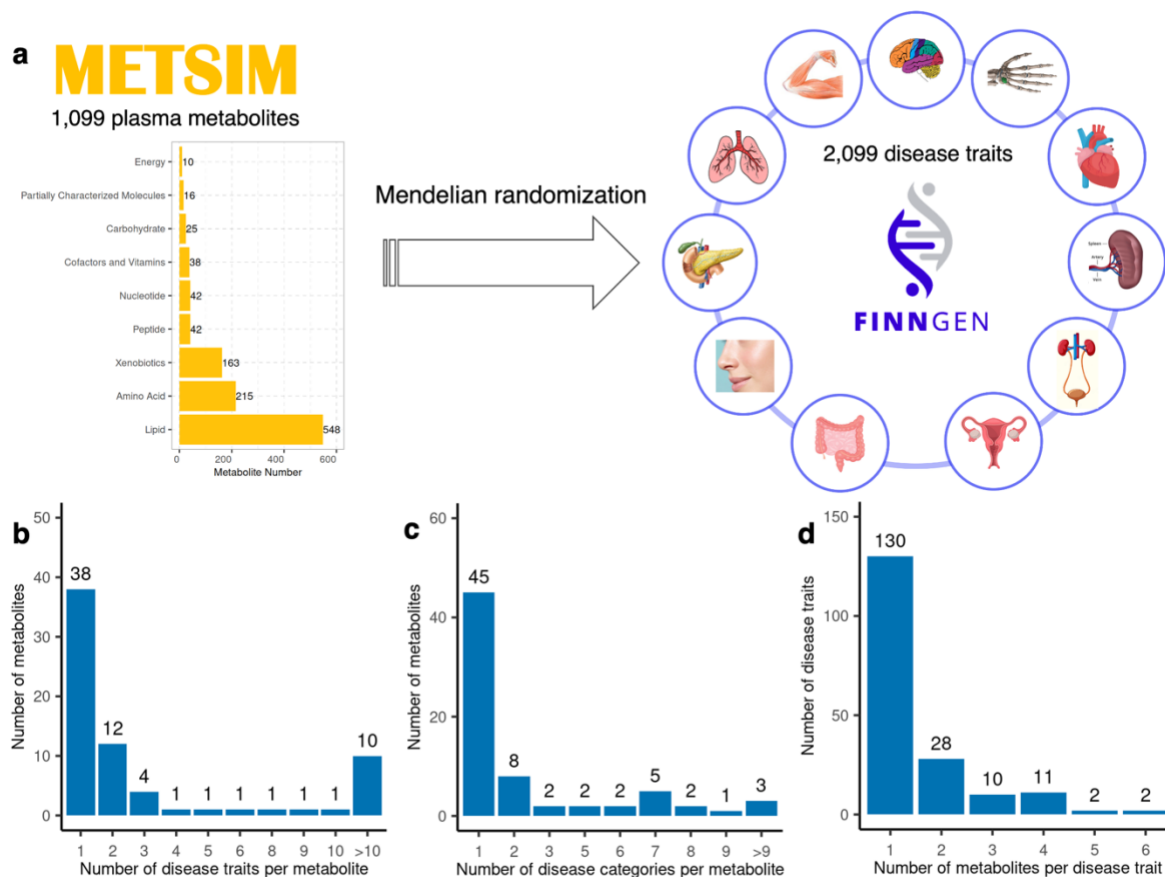
## 539 Reference

- 540 1 Wishart, D. S. Metabolomics for Investigating Physiological and  
541 Pathophysiological Processes. *Physiol Rev* **99**, 1819-1875 (2019).
- 542 2 Ahola-Olli, A. V. *et al.* Circulating metabolites and the risk of type 2 diabetes: a  
543 prospective study of 11,896 young adults from four Finnish cohorts. *Diabetologia*  
544 **62**, 2298-2309 (2019).
- 545 3 Reznik, E. *et al.* A Landscape of Metabolic Variation across Tumor Types. *Cell Syst*  
546 **6**, 301-313.e303 (2018).
- 547 4 Wang, T. J. *et al.* Metabolite profiles and the risk of developing diabetes. *Nat Med*  
548 **17**, 448-453 (2011).
- 549 5 Wu, Q. *et al.* Prediction of Metabolic Disorders Using NMR-Based Metabolomics:  
550 The Shanghai Changfeng Study. *Phenomics* **1**, 186-198 (2021).
- 551 6 Surendran, P. *et al.* Rare and common genetic determinants of metabolic  
552 individuality and their effects on human health. *Nat Med* **28**, 2321-2332 (2022).
- 553 7 Bartel, J. *et al.* The Human Blood Metabolome-Transcriptome Interface. *PLoS*  
554 *Genet* **11**, e1005274 (2015).
- 555 8 Yin, X. *et al.* Genome-wide association studies of metabolites in Finnish men  
556 identify disease-relevant loci. *Nat Commun* **13**, 1644 (2022).
- 557 9 Yin, X. *et al.* Integrating transcriptomics, metabolomics, and GWAS helps reveal  
558 molecular mechanisms for metabolite levels and disease risk. *Am J Hum Genet*  
559 **109**, 1727-1741 (2022).
- 560 10 Richmond, R. C. & Davey Smith, G. Mendelian Randomization: Concepts and  
561 Scope. *Cold Spring Harb Perspect Med* **12** (2022).
- 562 11 Zhao, Q., Wang, J., Hemani, G., Bowden, J. & Small, D. S. Statistical inference in  
563 two-sample summary-data Mendelian randomization using robust adjusted  
564 profile score. *The Annals of Statistics* **48**, 1742-1769, 1728 (2020).
- 565 12 Burgess, S. & Thompson, S. G. Multivariable Mendelian randomization: the use of  
566 pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol* **181**, 251-  
567 260 (2015).
- 568 13 Wang, J. *et al.* Causal inference for heritable phenotypic risk factors using  
569 heterogeneous genetic instruments. *PLoS Genet* **17**, e1009575 (2021).
- 570 14 Sun, Y., Lu, Y. K., Gao, H. Y. & Yan, Y. X. Effect of Metabolite Levels on Type 2  
571 Diabetes Mellitus and Glycemic Traits: A Mendelian Randomization Study. *J Clin*  
572 *Endocrinol Metab* **106**, 3439-3447 (2021).
- 573 15 Qian, L. *et al.* Genetically Determined Levels of Serum Metabolites and Risk of  
574 Neuroticism: A Mendelian Randomization Study. *Int J Neuropsychopharmacol* **24**,  
575 32-39 (2021).

- 576 16 Lord, J. *et al.* Mendelian randomization identifies blood metabolites previously  
577 linked to midlife cognition as causal candidates in Alzheimer's disease. *Proc Natl*  
578 *Acad Sci U S A* **118** (2021).
- 579 17 Qin, Y. *et al.* Genome-wide association and Mendelian randomization analysis  
580 prioritizes bioactive metabolites with putative causal effects on common  
581 diseases. *medRxiv* (2020).
- 582 18 Chen, Y. *et al.* Genomic atlas of the plasma metabolome prioritizes metabolites  
583 implicated in human diseases. *Nat Genet* **55**, 44-53 (2023).
- 584 19 Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped  
585 isolated population. *Nature* **613**, 508-518 (2023).
- 586 20 Davies, N. M., Holmes, M. V. & Davey Smith, G. Reading Mendelian randomisation  
587 studies: a guide, glossary, and checklist for clinicians. *Bmj* **362**, k601 (2018).
- 588 21 Shu, H. *et al.* Emerging Roles of Ceramide in Cardiovascular Diseases. *Aging Dis*  
589 **13**, 232-245 (2022).
- 590 22 Lotta, L. A. *et al.* Genetic Predisposition to an Impaired Metabolism of the  
591 Branched-Chain Amino Acids and Risk of Type 2 Diabetes: A Mendelian  
592 Randomisation Analysis. *PLoS Med* **13**, e1002179 (2016).
- 593 23 Ferkingstad, E. *et al.* Large-scale integration of the plasma proteome with  
594 genetics and disease. *Nat Genet* **53**, 1712-1721 (2021).
- 595 24 Ngo, D. *et al.* Proteomic profiling reveals biomarkers and pathways in type 2  
596 diabetes risk. *JCI Insight* **6** (2021).
- 597 25 Huang, S. Y. *et al.* Investigating Causal Relations Between Circulating Metabolites  
598 and Alzheimer's Disease: A Mendelian Randomization Study. *J Alzheimers Dis* **87**,  
599 463-477 (2022).
- 600 26 Zhuang, Z. *et al.* Causal relationships between gut metabolites and Alzheimer's  
601 disease: a bidirectional Mendelian randomization study. *Neurobiol Aging* **100**,  
602 119.e115-119.e118 (2021).
- 603 27 Chen, H. *et al.* Assessing Causal Relationship Between Human Blood Metabolites  
604 and Five Neurodegenerative Diseases With GWAS Summary Statistics. *Front*  
605 *Neurosci* **15**, 680104 (2021).
- 606 28 Knuplez, E. & Marsche, G. An Updated Review of Pro- and Anti-Inflammatory  
607 Properties of Plasma Lysophosphatidylcholines in the Vascular System. *Int J Mol*  
608 *Sci* **21** (2020).
- 609 29 Kinney, J. W. *et al.* Inflammation as a central mechanism in Alzheimer's disease.  
610 *Alzheimers Dement (N Y)* **4**, 575-590 (2018).
- 611 30 Semba, R. D. Perspective: The Potential Role of Circulating  
612 Lysophosphatidylcholine in Neuroprotection against Alzheimer Disease. *Adv*  
613 *Nutr* **11**, 760-772 (2020).
- 614 31 Hammouda, S. *et al.* Genetic variants in FADS1 and ELOVL2 increase level of  
615 arachidonic acid and the risk of Alzheimer's disease in the Tunisian population.  
616 *Prostaglandins Leukot Essent Fatty Acids* **160**, 102159 (2020).
- 617 32 Kovesdy, C. P. Epidemiology of chronic kidney disease: an update 2022. *Kidney*  
618 *Int Suppl (2011)* **12**, 7-11 (2022).
- 619 33 Niwa, T. Role of indoxyl sulfate in the progression of chronic kidney disease and  
620 cardiovascular disease: experimental and clinical effects of oral sorbent AST-120.  
621 *Ther Apher Dial* **15**, 120-124 (2011).
- 622 34 Leite, M. I. *et al.* Myasthenia gravis and neuromyelitis optica spectrum disorder: a  
623 multicenter study of 16 patients. *Neurology* **78**, 1601-1607 (2012).

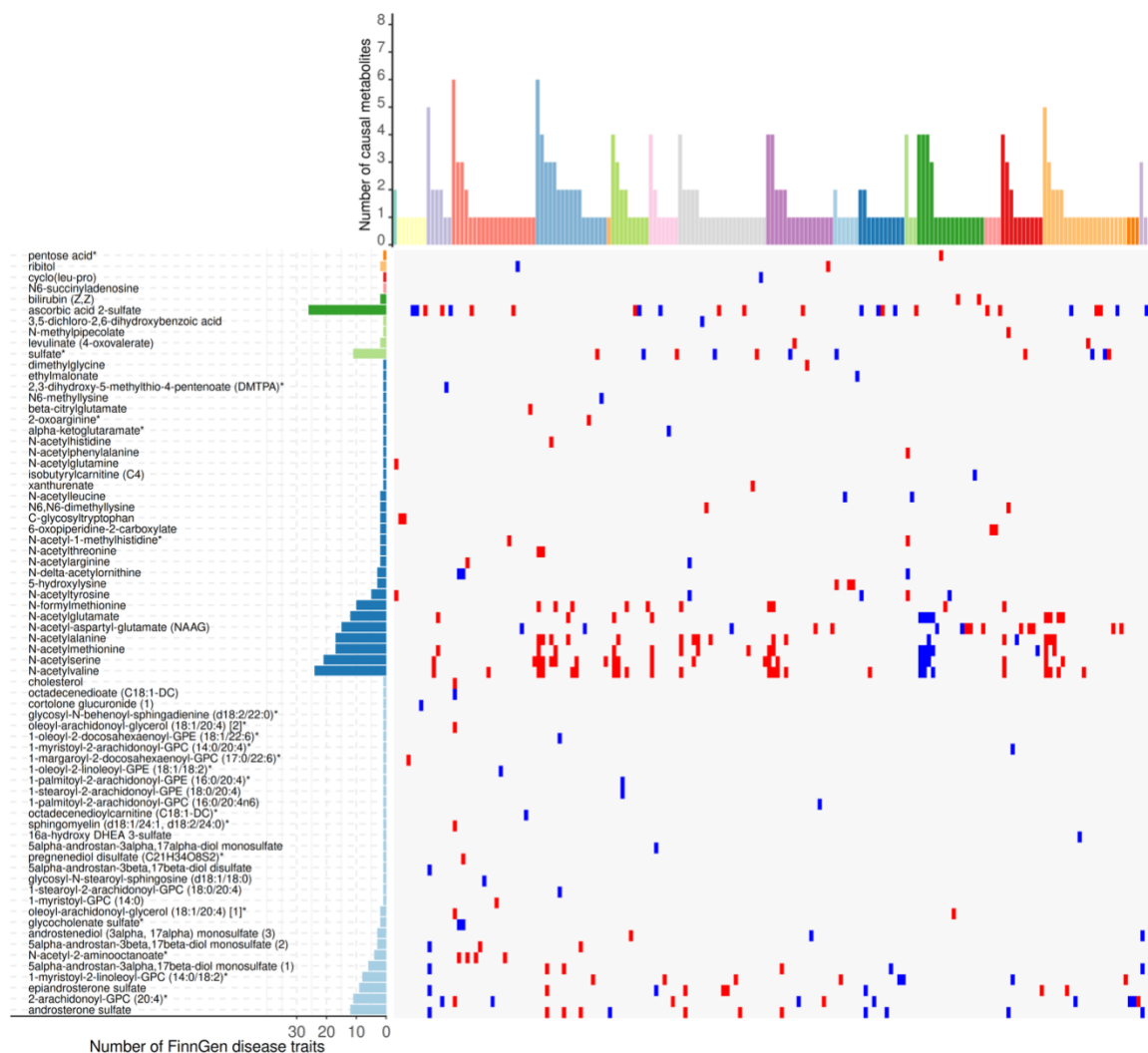
- 624 35 Li, X., Sundquist, J. & Sundquist, K. Subsequent risks of Parkinson disease in  
625 patients with autoimmune and related disorders: a nationwide epidemiological  
626 study from Sweden. *Neurodegener Dis* **10**, 277-284 (2012).
- 627 36 Li, Y. & Schellhorn, H. E. New developments and novel therapeutic perspectives  
628 for vitamin C. *J Nutr* **137**, 2171-2184 (2007).
- 629 37 Gatto, G. J., Jr., Boyne, M. T., 2nd, Kelleher, N. L. & Walsh, C. T. Biosynthesis of  
630 pipercolic acid by RapL, a lysine cyclodeaminase encoded in the rapamycin gene  
631 cluster. *J Am Chem Soc* **128**, 3838-3847 (2006).
- 632 38 Alonso, A. *et al.* Serum Metabolomics and Incidence of Atrial Fibrillation (from  
633 the Atherosclerosis Risk in Communities Study). *Am J Cardiol* **123**, 1955-1961  
634 (2019).
- 635 39 Laakso, M. *et al.* The Metabolic Syndrome in Men study: a resource for studies of  
636 metabolic and cardiovascular diseases. *J Lipid Res* **58**, 481-493 (2017).
- 637 40 Lim, E. T. *et al.* Distribution and medical impact of loss-of-function variants in the  
638 Finnish founder population. *PLoS Genet* **10**, e1004494 (2014).
- 639 41 Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample  
640 relatedness in large-scale genetic association studies. *Nat Genet* **50**, 1335-1341  
641 (2018).
- 642 42 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-  
643 based linkage analyses. *Am J Hum Genet* **81**, 559-575 (2007).
- 644 43 Mbatchou, J. *et al.* Computationally efficient whole-genome regression for  
645 quantitative and binary traits. *Nat Genet* **53**, 1097-1103 (2021).
- 646
- 647

648 **Figure 1: Summary of the 282 significant causal effects of 70 metabolites on 183**  
 649 **disease traits.** **a**, the overall design of univariable Mendelian randomization to test  
 650 causal effects of 1,099 metabolites on 2,099 disease traits; **b**, distribution of metabolites  
 651 by the number of disease traits that they showed significant causal effects on; **c**,  
 652 distribution of metabolites by the number of disease categories that they showed  
 653 significant causal effects on; **d**, distribution of disease traits by the number of their  
 654 associated causal metabolites.

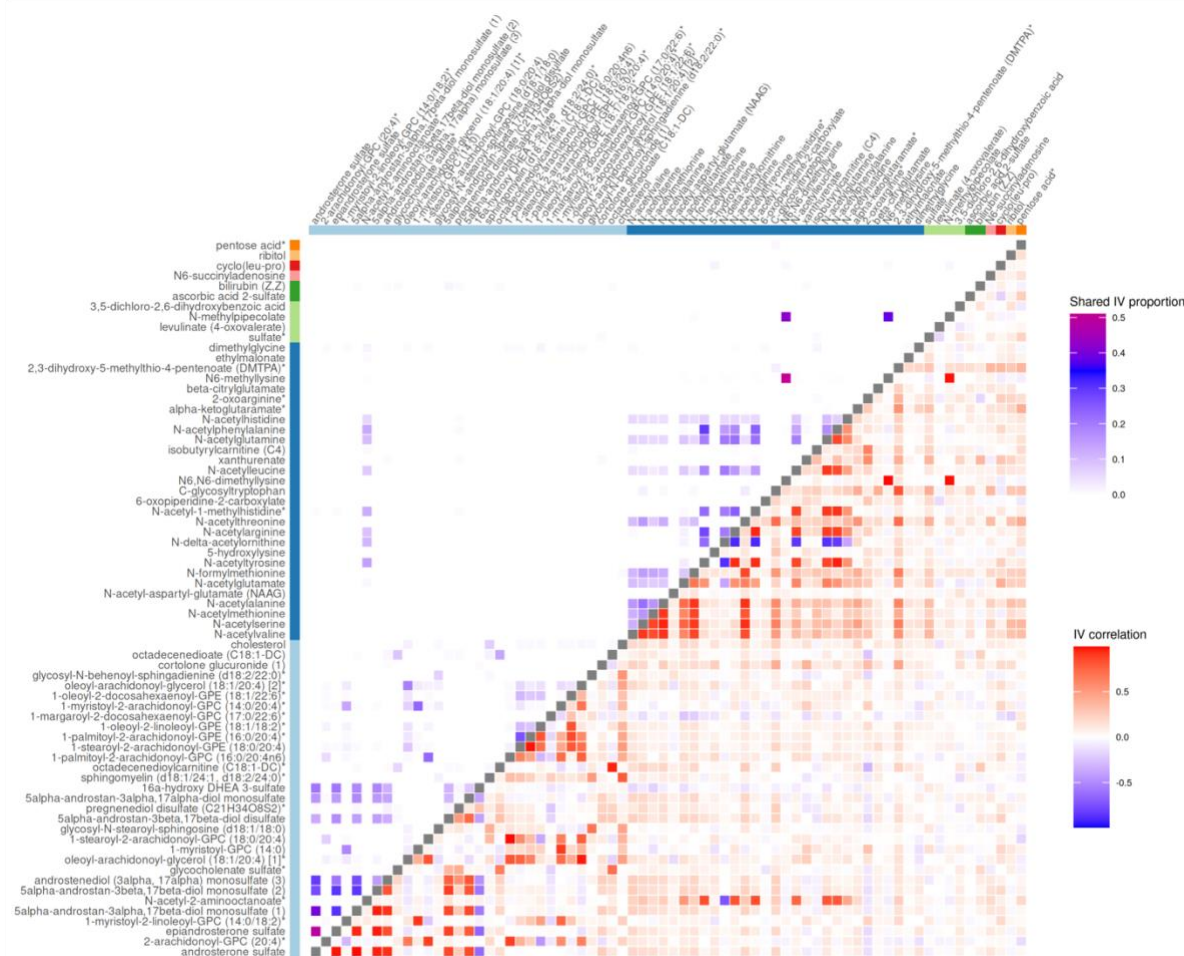


655  
656

657 **Figure 2: Heat map of the 282 potential causal effects of 70 metabolites on 183**  
658 **FinnGen disease traits.** The x-axis denotes the 183 disease traits of 20 colored  
659 categories (from left to right): light paleturquoise (n=1; alcohol related diseases), light  
660 wheat (7; congenital malformations, deformations and chromosomal abnormalities),  
661 light steel blue (6; diseases of the blood and blood-forming organs), salmon (20; diseases  
662 of the circulatory system), sky blue (17; diseases of the digestive system), dark sandy  
663 brown (1; diseases of the ear and mastoid process), dark olive green (9; diseases of the  
664 eye and adnexa), light thistle (7; diseases of the genitourinary system), gray (21; diseases  
665 of the musculoskeletal system and connective tissue), orchid (16; diseases of the nervous  
666 system), light sky blue (6; diseases of the respiratory system), dodger blue (11; diseases  
667 of the skin and subcutaneous tissue), dark sea green (3; drug purchase endpoints), forest  
668 green (16; endocrine, nutritional and metabolic diseases), light pink (4; infectious and  
669 parasitic diseases), fire brick (10; mental and behavioral disorders), sandy brown (20;  
670 neoplasms), dark orange (3; neurological diseases), medium thistle (4; pregnancy,  
671 childbirth and the puerperium), and medium purple (1; rheuma endpoints). The y-axis  
672 denotes the 70 metabolites of eight colored biochemical classes (from bottom to top):  
673 light blue (n=31; lipids), dark blue (29; amino acids), light green (4; xenobiotics), dark  
674 green (2; cofactors and vitamins), pink (1; nucleotides), red (1; peptides), light orange (1;  
675 carbohydrates), and dark orange (1; partially characterized molecules). The bar plots  
676 show the number of FinnGen disease traits that each metabolite confers causal effects on  
677 (on the left) and the number of causal metabolites for each disease trait (on the top). The  
678 color of cells denotes the direction of potential causal effects (red for positive and blue  
679 for negative effects) of metabolites on disease traits.  
680



683 **Figure 3: IV sharing (upper left triangular heat map) and correlation (lower right**  
 684 **triangular heat map) for all pairs of the 70 metabolites.** The color bar on the x-axis  
 685 and y-axis denotes the biochemical classes of metabolites: light blue (lipids), dark blue  
 686 (amino acids), light green (xenobiotics), dark green (cofactors and vitamins), pink  
 687 (nucleotides), red (peptides), light orange (carbohydrates), and dark orange (partially  
 688 characterized molecules). In the upper left triangular heat map, each cell denotes the  
 689 proportion of IVs with metabolite association  $P \leq 10^{-5}$  shared between the pair of  
 690 metabolites. In the lower right triangular heat map, and each cell denotes the IV  
 691 correlation between the pair of metabolites. The diagonal cells are colored in dark gray  
 692 to distinguish the upper and lower triangular heat maps.  
 693



694  
695



696 **Figure 4: Mendelian randomization suggests two metabolic pathways for anxious**  
697 **personality disorder.** Genes implicated for the ratio of N6,N6-dimethyllysine and N-  
698 methylpipercolate and for androsterone sulfate are italicized.  
699

