

1 Multi-trait genome-wide association study in 34,394 Chinese
2 women reveals the genetic architecture of plasma metabolites
3 during pregnancy

4
5 Siyang Liu,^{1,3#} Jilong Yao,^{2#} Liang Lin,^{4#} Xianmei Lan,^{1,5#} Linlin Wu,^{2,6#} Nannan
6 Kong,⁴ Yuqing Deng,⁷ Yan Li,¹ Jiansheng Xie,² Huanhuan Zhu¹, Xiaoxia Wu,^{2,6,8}
7 Zilong Li,¹ Likuan Xiong,⁹ Yuan Wang¹, Jinghui Ren,¹⁰ Xuemei Qiu,⁴ Weihua
8 Zhao,¹¹ Ya Gao,¹ Yuanqing Chen,¹² Fengxia Su,¹ Yun Zhou,¹³ Weiqiao Rao,⁴ Jing
9 Zhang,¹⁴ Guixue Hou,⁴ Liping Huang,¹⁵ Linxuan Li^{1,5}, Xinhong Liu,¹⁶ Chao Nie,¹
10 Liqiong Luo,¹⁷ Zengyou Liu,¹⁸ Fang Chen,¹⁹ Shengmou Lin,²⁰ Lijian Zhao,⁴ Qingmei
11 Fu,²¹ Dan Jiang,⁴ Ye Yin,²² Xun Xu,^{1,23} Jian Wang,²² Huanming Yang,^{1,24} Rong
12 Wang,^{4,5} Jianmin Niu^{2*}, Fengxiang Wei,^{25*} Xin Jin,^{1,26,*} Siqu Liu,^{1,*}

13
14 1 BGI-Shenzhen, Shenzhen 518083, Guangdong, China
15 2 Shenzhen Maternity & Child Healthcare Hospital, Shenzhen, Guangdong, 518000, China
16 3 School of Public Health (Shenzhen), Sun Yat-sen University, Shenzhen, Guangdong
17 510006, China
18 4 BGI Genomics, BGI-Shenzhen, Shenzhen 518083, Guangdong, China
19 5 College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049,
20 China
21 6 Department of Obstetrics, Shenzhen Maternity & Child Healthcare Hospital, Shenzhen,
22 Guangdong, 518000, China
23 7 Peking University Shenzhen Hospital, Shenzhen 518035, Guangdong, China
24 8 Department of Obstetrics, Shenzhen Maternity & Child Healthcare Hospital, The First
25 School of Clinical Medicine, Southern Medical University, Shenzhen, Guangdong , 518000,
26 China
27 9 Baoan Women's and Children's Hospital, Jinan University, Shenzhen 518133, Guangdong,
28 China
29 10 Shenzhen People's Hospital, 2nd Clinical Medical College of Jinan University, Shenzhen
30 518020, Guangdong, China
31 11 Shenzhen Second People Hospital, Shenzhen 518035, Guangdong, China
32 12 Nanshan Medical Group Headquarters of Shenzhen, Shenzhen 518000, Guangdong, China
33 13 Luohu People's Hospital of ShenZhen, Shenzhen 518001, Guangdong, China
34 14 Shenzhen Nanshan Maternity&Child Healthcare Hospital, Shenzhen 518000, Guangdong,
35 China
36 15 Shenzhen Baoan District Shajing People's Hospital, Shenzhen 518104, Guangdong,
37 Chinas
38 16 Shenzhen Longhua District Central Hospital, Shenzhen 518110, Guangdong, China
39 17 The People's Hospital of Longhua-Shenzhen, Shenzhen 518109, Guangdong, China
40 18 Shenzhen Nanshan People's Hospital, Shenzhen 518052, Guangdong, China
41 19 MGI, BGI-Shenzhen, Shenzhen 518083, China
42 20 The University of Hong Kong - Shenzhen Hospital, Shenzhen 518038, Guangdong, China
43 21 Baoan People's Hospital of Shen Zhen, Shenzhen 518100, Guangdong, China
44 22 BGI, BGI-Shenzhen, Shenzhen 518083, China
45 23 Guangdong Provincial Key Laboratory of Genome Read and Write
46 24 Guangdong Provincial Academician Workstation of BGI Synthetic Genomics
47 25 Longgang District Maternity&Child Healthcare Hospital of Shenzhen City, Shenzhen
48 518172, Guangdong, China
49 26 School of Medicine, South China University of Technology, Guangzhou 51006,
50 Guangdong, China

51 # Those authors contributed equally
52 \$ Senior author
53 *Correspondence to Siqi Liu siqi.liu@genomics.cn, Xin Jin jinxin@genomics.cn, Fengxiang
54 Wei haowei727499@163.com and Jianmin Niu njianmin@163.com.
55

56 **Summary**

57 Metabolites are important indicators of individual health and can serve as crucial
58 targets for therapy. However, the genetic basis of many metabolites remains largely
59 unexplored, especially among underrepresented East Asians and during critical
60 periods such as pregnancy. In this study, we utilized genetic information obtained
61 from non-invasive prenatal testing to conduct a genome-wide association analysis of
62 84 metabolites, including 37 amino acids, 10 vitamins, 24 metal elements, and 13
63 hormones, among 34,394 Chinese pregnant women. Of these metabolites, 52 and 11
64 had not previously been studied in East Asians or globally. We identified 30 novel
65 metabolite-gene associations. We also observed substantial differences in the genetic
66 effects on hormones between pregnancy and non-pregnancy periods, suggesting effect
67 modifications in response to physiological changes. Furthermore, we uncovered
68 pervasive pleiotropic effects for 50.94% of the genetic associations among
69 metabolites, as well as between six metabolites and eight pregnancy biomarkers.
70 Using mendelian randomization, we identified potential causal relationships between
71 plasma folate and ischemic stroke, vitamin D3 and Graves' disease, copper and open-
72 angle glaucoma, and androstenedione and rheumatoid arthritis. These discoveries
73 provide invaluable genetic insights into human metabolism, laying the foundation for
74 future mechanistic studies and the development of new therapeutic targets,
75 particularly for underrepresented East Asians.
76 **Keywords:** metabolites, genetic basis, multi-trait genome-wide association study,
77 pregnancy, East Asians

78 **5617 words**

79 **Introduction**

80 Metabolites are small molecules that act as intermediate or end products during the
81 metabolic processes¹. Their concentrations are influenced by both genetic and
82 environmental factors^{2,3}. Metabolites can have a crucial role in disease etiology and
83 are often targeted in therapeutic interventions^{4,5}. Human genetics provides an
84 important approach to understanding the determinants of metabolite alteration and
85 assessing the role of metabolites in disease outcomes⁶. Genome-wide association
86 studies (GWAS) of metabolites have identified thousands of variants associated with
87 approximately 7,000 metabolites in blood plasma or urine, as reported in the latest
88 GWAS catalog (r2023-01-30)⁷. However, like many other traits, human genetic
89 studies of metabolites have been dominated by European populations⁸. Specifically,
90 less than 300 metabolites (less than 4.29% of all metabolites) have been studied in
91 East Asian populations, with generally small sample sizes (less than 2,000
92 individuals)⁹⁻¹⁶. Therefore, there is a significant gap in our understanding of the
93 genetic effects of metabolites in underrepresented East Asian populations.
94

95 Previous genetic studies of metabolites have not yet addressed the question of whether
96 and how genetic effects may vary during pregnancy, a special and critical period that
97 plays a pivotal role in human reproduction. Pregnancy involves a series of
98 physiological changes and metabolic adaptations that necessitate a balanced
99 metabolome for the health of both the mother and the fetus¹⁷. Observational
100 epidemiological studies have demonstrated that inadequate or excessive metabolite
101 concentrations can result in poor health outcomes or severe disease and adverse birth
102 outcomes. For example, excessive vitamin A levels may cause fetal malformation¹⁸, a
103 lack of folic acid (vitamin B9) was associated with a high risk of neural tube defects¹⁹,
104 a deficiency of selenium was related to an elevated risk of miscarriage, pre-eclampsia,
105 and growth restriction^{20,21}, and vitamin D insufficiency was related to fetal bone
106 growth retardation²². To investigate the causal relationship and molecular
107 mechanisms underlying these observational associations, genetic studies of plasma
108 metabolite levels during pregnancy are crucial. However, maternal metabolites during
109 pregnancy have not been explored using genome-wide association studies, and the
110 genetic basis of metabolite levels during pregnancy is largely unknown. It is pertinent
111 to investigate whether the genetic effects on metabolites, as estimated from genome-
112 wide association studies, may differ between pregnant and non-pregnant populations.
113 Moreover, it is important to investigate whether metabolites assayed during this
114 critical early period of human life may have a causal relationship with diseases that
115 manifest later in life. Such investigations will yield critical insights into human
116 metabolism, providing the foundation for downstream mechanistic studies and the
117 identification of new therapeutic targets.

118
119 Conducting large-scale genetic analyses, particularly in underrepresented populations,
120 remains a significant challenge. In a prior pioneering study, we developed
121 methodologies and protocols for robust genetic association analysis using non-
122 invasive prenatal testing (NIPT) data.²³ In this study, we conducted genetic analyses
123 of a set of 84 maternal metabolites (abbreviated as MM) during pregnancy, leveraging
124 genetic resources obtained from NIPT and metabolite measurements from tandem
125 mass spectrometry among 34,394 Chinese pregnant women. These 84 metabolites
126 included 37 amino acids, 5 hydrophilic vitamins, 5 hydrophobic vitamins, 13
127 hormones, and 24 trace elements and heavy metals. Of these metabolites, 52 have not
128 been studied in East Asians, and 11 have not been investigated globally before (Table
129 S1). The exact number of participants involved for each metabolite was detailed in
130 Table S2. We performed both single- and multi-trait genome-wide association
131 analyses to identify genetic signals that significantly contribute to plasma metabolite
132 levels and to estimate SNP heritability. We compared the genetic effects of the same
133 metabolite measurements estimated from both pregnancy participants and non-
134 pregnancy individuals. Additionally, we conducted wald ratio estimation and
135 colocalization analyses to identify pleiotropic effects among metabolites in the
136 context of the KEGG pathway and between metabolites and several pregnancy
137 phenotypes reported in a companion study. Finally, we applied mendelian
138 randomization to investigate the potential causal effects of pregnancy metabolites on
139 disease traits in East Asian populations. This study represents the largest scale
140 metabolite genome-wide association study in underrepresented East Asian
141 populations and has provided the first insight into the genetic basis of the plasma level
142 of 84 metabolites during pregnancy.

143 Results

144 Study design and phenotypic distribution of the 84 metabolites

145 The 34,394 pregnant participants were recruited during their routine obstetric
146 examinations in the city of Shenzhen, South China. The expense of the non-invasive
147 prenatal test (NIPT) is covered by the government and the insurance in the Shenzhen
148 city. When NIPT is provided, each participant was asked if they would like to partake
149 in the pregnancy nutrition program for screening of metabolites. We did not identify a
150 difference of ages between participants partaking in the nutrition programs and
151 those did not. The mean age of the participants was 30 years (standard deviation
152 [s.d.] 4.9), and the mean gestational week for blood sampling was 16 weeks plus 3
153 days (s.d. 4.1 weeks) (Table S2). Non-invasive prenatal tests (NIPT) were provided as
154 a routine prenatal screening test²⁴, while tandem mass spectrometry (MS/MS) assays
155 of 84 metabolites covering four categories were offered as an additional screening
156 option to monitor the metabolite status of the participants using the same tube of
157 blood plasma.

158
159 For the NIPT test, whole genome low-pass sequencing (0.1x - 0.3x) was carried out,
160 and we applied NIPT sequencing of the 34,394 participants using two sequencing
161 platforms to minimize potential technical artifacts due to low-pass sequencing.
162 Specifically, 18,091 of them were sequenced using the blackbird sequencing machine
163 (BB) and 16,303 were sequenced using the Seq500 sequencing machine (Seq500).
164 We performed variation calling, genotype imputation, and genome-wide association
165 tests for data from each platform before we move on to the meta-analysis and the
166 multi-trait association study. The genotype imputation was performed using an
167 adaptive hidden Markov model with a Chinese population reference panel, and we
168 achieved imputation accuracy of 0.758 and 0.892 for BB and Seq500 data,
169 respectively (see Methods).

170
171 In the MS/MS assay, due to limited plasma volume, we separated the samples for
172 different assays including amino acids, hormones, water-soluble vitamins (WSV), fat-
173 soluble vitamins (FSV), and metal elements. The plasma volume limited our
174 capability to examine each metabolite for all the 34,394 participants. The mean
175 effective sample size was 7595, 10973, 8225 and 12033 for amino acid, hormone,
176 vitamin and metal elements, respectively and was detailed in Table S2. The plasma
177 concentration of 84 metabolites followed a Gaussian distribution with the presence of
178 outliers in a few of the metal elements (**Figure 1**). The mean value of the metabolites
179 fell in the reference ranges except for 3MHis, Al, I, Ti, ALD, and vitamin D, which
180 may be due to specific physiological status during pregnancy (Table S2).

181
182 The metabolites tend to function in a pathway, and therefore we suspected that
183 metabolites that play a role together in a specific pathway would be correlated with
184 each other. The phenotypic correlation between the metabolites suggests that
185 metabolites that belong to the same category tend to be more phenotypically
186 correlated with each other than metabolites that belong to different categories (Figure
187 S1, Table S3). Strong phenotypic correlations were observed between Met and His,
188 among Phe, Ile/Leu, and Tyr/Val for the amino acids, between Cd and Hg, among Al,
189 Ga, and V for metal elements, and between A4 and T, and between T1 and T2 for
190 hormones. The phenotypic correlation was smaller for metabolites between different

191 categories, with the strongest correlation observed between VA and P_{Ser}/Sar
192 (Spearman's R ~ 0.41) and between VE and Car (Spearman's R ~ 0.42).
193

194 **Genome-wide association analysis identifies 53 metabolite-gene** 195 **associations**

196 Based on the abovementioned observation of the phenotypic correlation within the
197 same metabolite categories, we conducted a multi-trait genome-wide association
198 study for each metabolite by utilizing information obtained from GWAS of
199 metabolites from the same category. Firstly, we performed a regression analysis
200 using PLINK²⁵, to regress the quantile-transformed metabolite value on the genotype
201 dosage for BB and Seq500 sequencing data, respectively. The analysis was adjusted
202 for confounding factors, including the first three principal components, maternal age,
203 gestational week of non-invasive prenatal testing (NIPT), MS/MS testing, and the
204 inferred sex of the fetus. Detailed information can be found in Table S4.
205 Subsequently, we conducted a meta-analysis using METAL²⁶ to combine GWAS
206 summary statistics from BB and Seq500. Finally, we utilized the multi-trait
207 association test algorithm implemented in MTAG²⁷ to generate trait-specific effect
208 estimates for each SNP, using information from summary statistics of
209 metabolites within the same metabolite categories.
210

211 The genomic control lambda (GC lambda) for the 84 metabolite molecules ranged
212 from 0.9 to 1.05, indicating a negligible inflation of association statistics. In total, we
213 identified 53 genetic associations reaching genome-wide significance ($p < 5 \times 10^{-8}$) for
214 33 metabolites, including 14 amino acids, 7 elements, 6 hormones, and 7 vitamins
215 (**Figure 2**). Among the 53 loci, 23 were previously known to be associated with the
216 corresponding metabolite level, while 30 were novel discoveries. The largest
217 standardized effect for each of the four metabolite categories was observed at (1) the
218 *CPS1-ERBB4* locus (lead SNP rs75472842, $\beta_{\text{nor}} = -0.46$, where an additional copy of
219 the T allele decreases 60.92 $\mu\text{mol/L}$ glycine in maternal plasma); (2) the *CXorf47*
220 locus (lead SNP rs1804495, $\beta_{\text{nor}} = -0.34$, where an additional copy the A allele
221 decreases 43.75 $\mu\text{g/L}$ iodine); (3) the *ZNF468 locus* (lead SNP rs4801940, $\beta_{\text{nor}} = -0.20$,
222 where an additional copy of the T allele decreases 0.45 ng/mL 17OHP); and (4) the
223 *LINC00441 locus* (rs144131853, $\beta_{\text{nor}} = -0.32$, where an additional copy of the T allele
224 decreases 0.32 ng/mL of vitamin D3) (**Table 1**). The complete information for the 53
225 genetic loci can be found in Table S5. We present the Manhattan, quantile-quantile,
226 and locuszoom plots for the 53 metabolite-gene associations in Figure S2. Below, we
227 summarize the key findings for each metabolite category.

228 ***Amino acids***

229 We found 18 independent association signals for 14 amino acids (**Figure 2a, Table**
230 **1**). All signals were first discovered in the East Asian population, and 12 were
231 previously reported in the European population and demonstrated consistent effect
232 estimates according to the GWAS catalog or PhenoScanner (Table S5). We focused
233 on the interpretation of the six novel association signals found for six amino acids.
234 The first novel signal was found for the 3-Methylhistidine (3MHis), an amino acid
235 biomarker for muscle protein turnover at the *CDK12* locus (lead SNP rs11078912-A:
236 $\beta = 0.07$, 95% CI: 0.04 to 0.09, $P = 1.88 \times 10^{-8}$) (**Figure 2a, Table S5**). The
237 rs11078912 variant was an expression quantitative trait locus (eQTL) for *FBL20*,
238 which plays a role in protein metabolism²⁸ and was strongly associated with HDL,

239 eGFR, and diseases such as asthma and rheumatoid arthritis²⁹. However, this
240 association had not been previously reported, likely due to limited genetic studies
241 (only four small-scale GWAS studies were documented for 3MHis in the GWAS
242 catalog).

243
244 Two novel association signals were identified between the *AGA* locus (lead SNP
245 rs12645096-T: beta=-0.13, 95% CI: -0.11 to -0.16, $P=9.07\times 10^{-26}$), the *ASPG* locus
246 (lead SNP rs1744294-T: beta=-0.13, 95% CI: -0.10 to -0.17, $P=1.54\times 10^{-16}$), and
247 aspartic acid (Asp). rs12645096 is a strong eQTL for *AGA*, which encodes the
248 aspartylglucosaminidase in many tissues such as artery, thyroid, and lungs (p-value <
249 $1.7e-32$) in GTEx. Similarly, rs1744294 is also an eQTL for *ASPG* that encodes
250 asparaginase. Both aspartylglucosaminidase and asparaginase deficiencies can lead to
251 an increased Asn and Asp, which explained the genetic association observed in our
252 study. Those associations had not been previously reported due to limited genetic
253 studies on this metabolite^{29,30}. While Asp has been used for the treatment of fatigue
254 and improvement of athletic performance and muscle strength without good scientific
255 evidence supporting its use³¹, the identified genetic association may serve as
256 instrumental variables for causal inference of its utility when combined with data
257 from biobanks with genomic data.

258
259 The fourth novel genetic association was identified for homocysteine (Hcy) at the
260 intergenic *VSTM5-HEPFL1* locus (lead SNP rs16919711-A: beta=-0.08, 95% CI: -
261 0.12 to -0.05, $P=2.90\times 10^{-8}$). The rs16919711 variant was an eQTL for *PANX1* and
262 *GPR83* in cultured fibroblasts cells and the cerebellar hemisphere in the brain.
263 However, the summary statistics of the previous 14 GWAS on Hcy were not available
264 for download, and we were not able to evaluate the effect size of rs16919711 in other
265 studies. As Hcy is an important biomarker for cardiovascular and cerebrovascular
266 disease³², it will be worthwhile to replicate and validate the function of this novel
267 association signal.

268
269 Finally, we found a novel association between sarcosine (Sar) and the intergenic
270 *SLC29A2-NPAS4* locus (lead SNP rs11227478-A: beta=0.08, 95% CI: 0.05 to 0.11,
271 $P=2.08\times 10^{-8}$) and an association between Valine (Val) and the *AGXT2* locus (lead
272 SNP rs191495-T: beta=-0.05, 95% CI: -0.03 to -0.07, $P=1.69\times 10^{-8}$). rs11227478 is
273 an eQTL for *CTSF* in several tissues and has been reported to associate with height
274 and forced vital capacity²⁹. There was limited knowledge of rs191495 at the
275 mitochondrial aminotransferase *AGXT2*. Information from future functional studies is
276 critical for interpreting these observed genetic associations.

277

278 ***Vitamins***

279 We identified 12 loci that have a strong impact on the levels of seven vitamins in
280 maternal plasma, six of which are known associations and six of which are novel
281 associations (**Figure 1b**). We have confirmed the previously known association
282 between the 5-methyltetrahydrofolate level (MF5) and the *NPPB-RNU5E-1* locus
283 (lead SNP rs75014290, $P=9.49\times 10^{-15}$), which is located near *MTHFR*, the gene that
284 encodes the methylenetetrahydrofolate reductase. We also identified a new
285 association between MF5 and variants within the *FOLR1-FOLR2* gene locus (lead
286 SNP rs35954619-A: beta=-0.09, 95% CI: -0.06 to -0.12, $P=3.00\times 10^{-9}$), an eQTL for
287 *FOLR1* which encodes the folate receptor 1 and folate receptor beta, and was

288 previously associated with the asymmetrical dimethylarginine levels related to heart
289 rate variability³³. On average, individuals with an additional T and A alleles for
290 rs75014290 and rs35954619, respectively, exhibit 2.78 ng/mL and 3.19 ng/ml lower
291 5-methyltetrahydrofolate in maternal plasma during pregnancy (see **Table 1**).

292
293 We have also identified a new association between vitamin B2 (VB2) and the
294 *SLC52A3* gene (lead SNP rs6140149-A: beta=-0.08, 95% CI: -0.05 to -0.10, $P=$
295 1.32×10^{-10}), which is known to play a role in water-soluble vitamin metabolism³⁴, as
296 well as a new association between vitamin B5 (VB5) and the *TCF23-SLC5A6* locus
297 (lead SNP rs1659696-A: beta=0.10, 95% CI: 0.07 to 0.13, $P= 3.92 \times 10^{-13}$), an eQTL
298 for *ATRAID*, which plays a role in cell cycle arrest in all-trans-retinoic acid signal
299 pathway²⁸. The discoveries of the three new genetic associations with the three water-
300 soluble vitamins, despite their strong biological relevance, were due to a lack of
301 previous genome-wide association studies (Table S1).

302
303 As for the fat-soluble vitamins, we identified the known associations between 25-
304 Hydroxyvitamin D3 level (VD3) and three loci *GC* (rs1352846, $P= 6.61 \times 10^{-33}$)
305 *PDE3B* (rs10766189, $P= 1.95 \times 10^{-20}$), and *DHCR7* (rs12789751, $P= 2.67 \times 10^{-15}$), as
306 well as two known associations for vitamin E (VE) at the *APOA5* ($P= 1.39 \times 10^{-8}$) and
307 vitamin K (VK) at the *LOC102724279-CYP4F2* locus ($P= 1.18 \times 10^{-37}$). Additionally,
308 we have identified a new association for vitamin A (VA) at the *GCKR* locus (lead
309 SNP rs1260333-A: beta=0.07, 95% CI: 0.05 to 0.10, $P= 5.31 \times 10^{-9}$); a new
310 association for VD3 at the *LINC00441* locus (lead SNP rs144131853-A: beta=-0.04,
311 95% CI: 0.05 to 0.09, $P=1.60 \times 10^{-9}$); and two new associations for VE at the *RNF215*
312 (lead SNP rs61583714-T: beta=0.14, 95% CI: 0.12 to 0.17, $P= 2.31 \times 10^{-28}$) and *GGH*
313 (lead SNP rs117494536-A: beta=0.12, 95% CI: 0.10 to 0.15, $P= 8.16 \times 10^{-20}$) locus.
314 The *LINC00441*, *RNF215*, and the *GGH* loci were eQTL for *RBI* in adipose tissue,
315 *SEC14L3* in Thyroid, and *GGH* in the whole blood, suggesting biological relevance.
316

317 **Metal elements**

318 We identified eleven genetic associations for seven different metal elements,
319 including three previously reported and eight novel signals (**Figure 2b**). The three
320 known loci were the *CP* locus for copper level (Cu) (lead SNP rs17838831: $P=$
321 1.46×10^{-30}), the *TMPRSS6* locus for iron level (Fe) (lead SNP rs877908: $P= 2.41 \times 10^{-$
322 26), and the *BHMT2* locus for selenium level (Se) (lead SNP rs2909855: $P= 2.00 \times 10^{-$
323 36).

324
325 The eight novel associations included the *GIGYF2* locus, an eQTL for *C2orf82* for the
326 barium level (Ba) through lead SNP rs13424351-A (beta=-0.10, 95% CI: -0.07 to -
327 0.12, $P= 4.32 \times 10^{-13}$); the *CCDC170* locus (lead SNP rs9397433-A: beta=-0.07, 95%
328 CI: -0.05 to -0.09, $P= 1.60 \times 10^{-9}$) and the *LOC401312-LOC541472* locus, an eQTL
329 for the *IL6* antisense RNA 1 (*IL6-AS1*) (lead SNP rs58879058-T: beta=0.07, 95% CI:
330 0.05 to 0.10, $P= 2.00 \times 10^{-10}$) for copper level (Cu); the *SERPINA7* locus associated
331 with iodine level (I) through lead SNP rs1804495-A, which is a missense variant (
332 beta=-0.27, 95% CI: -0.25 to -0.30, $P= 2.13 \times 10^{-106}$); the *AGMAT* locus (lead SNP
333 rs6679229-A: beta=-0.06, 95% CI: -0.04 to -0.08, $P= 2.00 \times 10^{-9}$) and the *BORCS7-*
334 *ASMT* locus, an eQTL for *AS3MT* (lead SNP rs12764049-A: beta=-0.06, 95% CI: -
335 0.04 to -0.08, $P= 2.80 \times 10^{-9}$) were associated with the magnesium level (Mg); and
336 lastly the *EPRS* locus, an eQTL for *EPRS* (lead SNP rs12764049-T: beta=0.07, 95%

337 CI: 0.05 to 0.10, $P= 1.71 \times 10^{-9}$) and the *MORC4* locus, an eQTL for *PRPS1* (lead
338 SNP rs12764049-T: beta=0.16, 95% CI: 0.13 to 0.18, $P= 3.40 \times 10^{-39}$) were
339 associated with strontium level (Sr) in maternal plasma.

340

341 The discovery of the strong association between the missense variant rs1804495 at
342 *SERPINA7* and iodine level is noteworthy since no GWAS had been performed
343 previously for iodine. *SERPINA7* encodes the major thyroid hormone transport
344 protein, TBG, in serum. We found that one additional copy of the A allele decreases
345 11.12 ug/L iodine in maternal plasma, providing evidence for the regulation of iodine
346 metabolism through a common missense genetic polymorphism present in TBG.
347 Lastly, the remaining seven novel genetic associations displayed specific functions
348 and warrant further studies for biological validation.

349

350 **Hormone**

351 We identified 12 loci that have a genetic influence on the plasma levels of six
352 hormones, as presented in **Figure 2d**. Those included the known association between
353 the *CYP3A7-CYP3A51P* locus and the estrone (E1) (lead SNP rs4646462, $P=$
354 7.20×10^{-20}) and the association between the *BORCS7-ASMT* locus and the 17-
355 hydroxyprogesterone (17OHP) measurement (lead SNP rs10883790, $P= 1.45 \times 10^{-17}$).

356

357 We also identified ten novel signals. For instance, the *CYP2C18* locus was found to
358 associate with the 11-Deoxycorticosterone (DOC) (lead SNP rs12764049-A: beta=-
359 0.12, 95% CI: -0.15 to -0.10, $P= 6.74 \times 10^{-22}$). The lead SNP rs1926711 is an eQTL of
360 *CYP2C19*, a monooxygenase that catalyzes many reactions involved in the synthesis
361 of cholesterol, steroids, and other lipids and in drug metabolism.

362

363 In addition to the well-known association between the *CYP3A7-CYP3A51P* locus and
364 E1, we found four novel signals for E1. These signals included the *SULT1B1-*
365 *SULT1E1* locus (lead SNP rs1220712-T: beta=-0.08, 95% CI: 0.06 to 0.10, $P=$
366 2.11×10^{-13}), the *OASL-P2RX7* locus (lead SNP rs2708101-T: beta=0.07, 95% CI:
367 0.05 to 0.09, $P= 1.84 \times 10^{-9}$), the *LINC00379-MIR17HG* locus (lead SNP rs74241688-
368 A: beta=0.08, 95% CI: 0.06 to 0.11, $P= 3.27 \times 10^{-9}$) and the *VCX3A-PUDP* locus
369 (lead SNP rs6639741-A: beta=0.07, 95% CI: 0.05 to 0.09, $P= 3.81 \times 10^{-9}$).

370

371 Furthermore, we discovered a novel signal for Estradiol (E2) levels within *STS* (lead
372 SNP rs802900-: beta=0.06, 95% CI: 0.04 to 0.09, $P= 3.78 \times 10^{-8}$). Similar to the
373 *VCX3A-PUDP* locus for E1, this signal is also an eQTL for *PUDP*, which plays a role
374 in nucleotide salvage.

375

376 Lastly, we found three novel signals, including two for 11-Deoxycortisol (11DOC) at
377 the *CYP11B2-LOC100133669* (lead SNP rs143728108-T: beta=0.07, 95% CI: 0.04
378 to 0.09, $P= 1.88 \times 10^{-8}$) and the *FDXI* (lead SNP rs1073527-T: beta=0.08, 95% CI:
379 0.06 to 0.10, $P= 1.47 \times 10^{-10}$) locus, which plays roles in the steroid, vitamin D and
380 bile acid metabolism and one signal for 17 α -Hydroxyprogesterone (17OHP) at the
381 *ZNF468* locus (lead SNP rs4801940-T: beta=-0.10, 95% CI: -0.08 to -0.13, $P=$
382 9.49×10^{-15}).

383

384 **SNP heritability and genetic correlation**

385 To assess the heritability of the 84 metabolites, we employed the LD score regression
386 (LDSR) technique³⁵ to estimate SNP-based heritability. We then compared this
387 estimate to the heritability computed from summary statistics of GWAS from prior
388 investigations³⁶ (see Method). Among the 84 metabolites, 67 (78.6%) exhibited a SNP
389 heritability greater than zero (Figure S6, Table S6). The SNP-based heritability
390 estimates ranged from 0.4% to 22.2% with a median of 6.5% the 67 metabolites.
391 Notably, Ala ($h_{SNP}^2 = 22.2\%$), Hg ($h_{SNP}^2 = 17.7\%$), 17OHP ($h_{SNP}^2 = 10.1\%$), VD3
392 ($h_{SNP}^2 = 18.4\%$) and MF5 ($h_{SNP}^2 = 10.9\%$) demonstrated the highest SNP heritability
393 among the amino acid, metal elements, hormone, fat, and water vitamin categories,
394 respectively.

395 When comparing the SNP heritability of 15 metabolites from our study to the
396 publicly available GWAS summary statistics from prior investigations (Kettunen et
397 al., 2016 & Neale's lab), our study found that all three vitamins, VD3, VA, and VE, as
398 well as the metal element Mg and three amino acids, namely, Tyr, Ile, and Val,
399 exhibited higher SNP heritability (Table S6). However, for the remaining five amino
400 acids and two metal elements, our study found lower SNP heritability (Figure S7).

401

402 **Comparison of estimated genetic effects with a non-pregnancy Chinese** 403 **cohort**

404 All but eight of the 53 genetic loci identified in this study demonstrated the same
405 effect direction and reached nominal significance ($P < 0.05$) between the two different
406 sequencing technologies (Black Bird and BGI-Seq500) (**Figure S8**, Table S5). The
407 eight loci that did not reach nominal significance in one of the two sequencing
408 technologies displayed the same effect direction, likely due to lack of power. These
409 results provide support for the robustness of the GWAS hits.

410

411 To further explore the possibility of pregnancy-specific genetic associations, we
412 compared the effect size and the p-value for the 53 loci with an independent study of
413 1,553 non-pregnancy Chinese individuals (Female N=642, Male N=911) from the
414 BGI-Shenzhen multi-omics cohort³⁷. These 1,553 participants underwent whole-
415 genome sequencing and a total of 80 out of the 84 metabolites investigated in the
416 present study were also assayed for each individual using the same tandem MS
417 technology(see Methods). Notably in the BGI-Shenzhen multi-omics cohort, amino
418 acids, vitamins, and hormones were assayed using plasma, the same sampling
419 material as this study while the metal elements were assayed using whole blood. A
420 linear regression, adding individual sex and the first two PCs as the covariates, was
421 performed for each of the metabolites (see Method). A regression using only the
422 female individuals was also performed in the non-pregnancy cohort and the results
423 and the conclusions drawn below were consistent.

424

425 Interestingly, after excluding four signals for Hcy, Ba, I, and DOC that were not
426 assayed and for Hyl and Se that the lead SNPs were not present in the non-pregnancy
427 cohort, 25 loci discovered in the present study were not replicated in the non-
428 pregnancy cohort (47.2%, $P > 0.05$ or effect size were opposite)(Figure S9, Table S7).

429 The inconsistency in genetic effects between the pregnancy and non-pregnancy
430 cohorts varied among the four categories of metabolites. 5 out of the 16 signals in the
431 amino acid category (31.3%), 7 out of the 8 signals in the element category (87.5%),

432 all of the 10 signals in the hormone category (100%), and 2 out of 13 signals in
433 vitamin category (15.4%) demonstrated inconsistent genetic effects between the
434 pregnancy and non-pregnancy participants.

435
436 We summarized the reasons for the inconsistency in genetic effects for each category
437 as follows: (1) the low consistency for the element category was likely due to the
438 difference between the plasma and whole blood sampling, as differences in the mean
439 values of the elements between the two populations were observed (Table S7); (2) the
440 low consistency for hormones between the two studies may suggest pregnancy-
441 specific genetic effects, in accordance with the difference in the phenotypic
442 distribution between the two populations or may be due to a lack of power in the non-
443 pregnancy GWAS study; (3) for amino acids, the inconsistency of genetic effect for
444 the *ASPG* locus with Asp, the *SLC29A2-NPAS4* locus with Sar, and the *AGXT2* locus
445 with Val were likely due to the differences between the two cohorts, while the
446 inconsistency for the *KLKBI* locus with Arg and the *ASSI* locus with Cit was
447 probably due to lack of power, since the two association signals were previously
448 known. For vitamins, the overall phenotypic distribution between the two populations
449 was similar, except for MF5 and VB5, and therefore the rest of the association signals
450 tended to be consistent between the pregnancy and non-pregnancy cohorts.

451 **Extensive genetic pleiotropy among metabolites**

452 Metabolites were known to be interacting with each other, playing roles in certain
453 pathways, we inferred the existence of genetic pleiotropy for the metabolites
454 investigated in this study. Specifically, the 84 metabolites examined in this study were
455 involved in 42 KEGG pathways (Table S8)³⁸. The two largest pathways were the
456 aminoacyl-tRNA biosynthesis and steroid hormone biosynthesis pathways,
457 comprising 20 amino acids and 13 hormones in our study. We aim to explore whether
458 the genetic associations we identified suggest pleiotropy.

459
460 Among the 53 association signals, 27 (50.94%) affected two or more metabolite ($P <$
461 0.001 , Figure S10). By utilizing a wald ratio estimator to identify the potential causal
462 effect of one metabolite on another using a single SNP³⁹, we found 16 genetic
463 associations with a significant effect connecting 18 metabolites (Table S9). The
464 pathway analysis of these 18 metabolites revealed nine pathways with nominal
465 significance ($P < 0.05$) (Table S10). We visualized the pleiotropic effect in the
466 context of the KEGG pathways (**Figure 3**). For instance, in the cyanoamino acid
467 metabolism pathway, the *ASPG* locus (lead SNP rs1744297) suggested an effect from
468 Asp on Asn while the *AGA-LINC01098* locus (lead SNP rs12645096) locus affected
469 both the Asn and Asp. In addition, the *CPSI-ERBB4* locus (lead SNP rs10272406)
470 and *ZNF713* (lead SNP rs72933867) locus suggested an effect from Ser on Gly,
471 different from the direction suggested in the KEGG pathway (**Figure 3a**). In another
472 amino-acid-related aminoacyl-tRNA biosynthesis pathway, *KLKBI* (lead SNP
473 rs4253255) showed an effect from Arg on Met but not the other way around (**Figure**
474 **3b**). In the steroid hormone biosynthesis pathway, four gene loci, including
475 *LOC101928327-DIRC3-AS1* (lead SNP rs2162000), *EGR2-NRBF2* (lead SNP
476 rs10822140), *CYP3A5* (lead SNP rs1419745) and *JMJD1C* (lead SNP rs6479891)
477 mutually regulated A4-T-17OHP, E1-E2, E1-E2-CORT, and E1-E2-DHEA,
478 respectively (**Figure 3c**). For the ubiquinone and other terpenoid-quinone
479 biosynthesis pathway that involved vitamins, the *LOC102724279-CYP4F2* locus (lead

480 SNP rs12462273 and rs62107763) and *APOA5* locus (lead SNP rs3741298) affected
481 VK and VE simultaneously (**Figure 3d**).

482

483 The KEGG pathways lack information on interactions between trace elements and
484 between metabolites of different categories. Nonetheless, we found that the *CXorf57*
485 (rs72618342), *SERPINA7* (rs1804495), and *MORC4* (rs6616617) loci affected I and
486 Sr levels in the opposite direction (Figure S10). Moreover, the *EGR2-NRBF2*
487 (rs10822140) and *CYP3A5* (rs1419745) loci displayed bi-directional effects between
488 any two hormones in E2, E1, CORT, and Ser (Table S9). Lastly, the *GGH*
489 (rs72658350) locus suggested an effect of VE on Glu level but not the other way
490 around (Table S9, Figure S10).

491 **Causal maternal metabolites for pregnancy phenotypes and human** 492 **complex traits**

493 To investigate the newly discovered genetic associations with metabolites, we
494 performed a colocalization analysis for each of the 53 significant loci, using around
495 100 pregnancy phenotypes from a companion study (Zhu et al., manuscript in
496 submission). Furthermore, we conducted a bi-directional two-sample Mendelian
497 randomization (MR) analysis between the metabolites and 120 phenotypes from the
498 Biobank Japan Project (BBJ), released in the IEU open GWAS project⁴⁰ (see
499 Methods).

500

501 The results of the colocalization analysis revealed six loci that demonstrated a strong
502 signal of colocalization between maternal metabolites and pregnancy biomarkers
503 (posterior probabilities of H4, PPH4 > 0.5) (Figure S11). These included shared
504 pleiotropic effects between maternal Mg and serum creatinine (CR) and uric acid
505 (UA) at the *AGMAT* locus (Figure S11a), vitamin VA and serum creatinine (CR) and
506 pre-albumin (PA) at the *GCKR* locus (Figure S11b), E1, E2 and the mean platelet
507 volume (MPV) at the *EGFR-NRBF2-JMJD1C* locus (Figure S11c), 3MHis and urine
508 glucose (GLU_U) at the *FBXL20-CDK12* locus (Figure S11d), Fe and mean
509 corpuscular hemoglobin (MCH) at the *TST-TMPRSS6* locus (Figure S11e), and I and
510 free thyroxine (FT4) at the *SERPINA7* locus (Figure S11f). These findings contribute
511 to new knowledge of pleiotropy between maternal metabolites and biomarker traits
512 during pregnancy.

513

514 In the mendelian randomization analysis, after excluding heterogeneity, horizontal
515 pleiotropy, and reverse causality (see Methods), we identified significant causal
516 relationships between seven metabolites and 15 human traits and diseases from BBJ
517 ($P < 0.01$) (**Figure 4**). We found that genetically predicted MF5 was associated with
518 increased high-density lipoprotein (HDL) (beta=0.10, 95% CI: 0.04 to 0.16, $P =$
519 8.56×10^{-4}) and a reduced risk of ischemic stroke (OR=0.84, 95% CI: 0.73 to 0.95, $P =$
520 7.07×10^{-3}). The discovery of the relation between MF5 and these two traits verified
521 previous MR study and randomized controlled trials (RCT) in the European
522 population^{41,42}. Although MF5 protects against ischemic stroke, no causal effect of
523 MF5 on cardiovascular disease such as coronary artery disease was identified in the
524 BBJ ($P = 0.62$), consistent with the biological function of folate in the nucleic acid
525 synthesis and DNA repair, and medical evidence of folate in nervous system
526 development⁴³.

527

528 Furthermore, the MR analysis suggested that genetically predicted VD3 was
529 associated with an increased risk of Graves' disease (GD) (OR=1.40, 95% CI: 1.12 to
530 1.77, $P= 3.41 \times 10^{-3}$). This differs from observational studies, which suggested that
531 patients with GD were more likely to be deficient in VD compared to the controls,
532 based on a small sample size of less than 100⁴⁴. However, RCTs did not prove that
533 VD supplementation improves the course of GD⁴⁵. The most recent RCT observed a
534 trend for increased risk of treatment failure and relapse of hyperthyroidism with VD
535 supplementation⁴⁶. An MR analysis in the UK Biobank (UKBB) Europeans suggests
536 no significant associations between VD3 and hyperthyroidism⁴⁷. Therefore, to
537 confirm the effect of VD3 on GD observed in this study, further replication and
538 mechanism validation are required.

539
540 Regarding the metal elements, we found that genetically predicted plasma Cu was
541 associated with an increased risk of open-angle glaucoma (OR=1.51, 95% CI: 1.23 to
542 1.84, $P= 7.21 \times 10^{-5}$). As for the hormones, genetically predicted 11DOC was
543 associated with an increased risk of smoking initiation, consistent with previous
544 observational evidence⁴⁸, and a larger fractional shortening (FS, the size of the
545 ventricle at the end of systole and diastole) derived from the transthoracic
546 echocardiogram (TTE) data. In addition, A4 was associated with a higher level of
547 eGFR, total bilirubin, blood sugar, lower level of lactate dehydrogenase, urea nitrogen,
548 and a decreased risk of rheumatoid arthritis- consistent with reported observational
549 analysis⁴⁹. Lastly, E1 was associated with a lower level of chloride and alumin, which
550 was not reported previously.

551 Discussion

552 We identified a total of 53 genetic associations, including 30 novel genetic
553 associations for 19 metabolites. The consistency of the effect sizes of the associations
554 between the two sequencing technologies demonstrated the robustness of our
555 findings. Notably, our comparison of genetic associations with an independent cohort
556 of non-pregnancy individuals revealed substantial differences in genetic effects on
557 metabolites, especially those of the hormone category. These findings provide
558 valuable insights into human genetics, demonstrating that genetic effects can change
559 or be modified through interactions with internal and external environments, despite
560 DNA stability throughout life. Furthermore, the study also highlights the importance
561 of mechanistic studies to obtain an advanced understanding of the biological
562 mechanisms of pregnancy and reproduction.

563
564 We identified pervasive pleiotropic effects in 27 of the 53 metabolite-associated loci
565 (50.94%), which were enriched in nine KEGG pathways, suggesting genetic
566 determination of metabolic flux. Our study emphasizes the importance of multi-trait
567 association analysis rather than analyzing metabolites independently. We also
568 identified novel pleiotropic effects between metabolites of different categories,
569 including the shared effect between I and Sr, between three hormones (E2, E1,
570 CORT) and Ser, and between VE and Glu. Furthermore, analyzing the shared genetic
571 effects between metabolites and a hundred pregnancy phenotypes revealed novel
572 pleiotropy between six metabolites and eight pregnancy biomarkers, unraveling new
573 and complex functions of the metabolites. In our Mendelian randomization analysis of
574 metabolites and 120 phenotypes, we identified potential causal relationships between

575 plasma folate and ischemic stroke, vitamin D3 and Graves' disease, copper and open-
576 angle glaucoma, and androstenedione and rheumatoid arthritis. These novel findings
577 provide essential information for the development of new therapeutic targets for
578 complex diseases.

579

580 Our study has a few limitations. Firstly, due to a lack of birth cohort studies with
581 genetic data in China and East Asians, we were not able to evaluate the potential
582 causal impact of several maternal metabolites on severe pregnancy diseases and
583 adverse birth outcomes reported in observational studies. However, the genetic
584 discoveries obtained from this study made it possible to answer this question in the
585 future, along with the effects in China to build large-scale birth cohorts⁵¹. Our
586 findings of changing genetic effects during pregnancy, especially in hormones, raise
587 caution for Mendelian randomization studies, which examined the effect of
588 genetically predicted life-long exposures on the outcomes⁵². This will not be a
589 problem for the metabolites that demonstrated consistent genetic effects between the
590 pregnancy and the non-pregnancy status. However, for metabolites that with altering
591 genetic effects such as hormones, how may the pregnancy-specific genetic effect
592 impact late-onset disease will require further investigation. Our report on the potential
593 causal effects of maternal metabolites on human traits and diseases will require more
594 biological validation and replication from independent studies. Lastly, although the
595 current studies represented the largest-scale metabolite genome-wide association
596 studies in East Asians, expanding the sample size and types of metabolites will
597 facilitate a more systematic understanding of biological pathways. Finally, the genetic
598 associations identified from this study are from a single time point in early pregnancy,
599 we are expecting that more metabolite changes will appear in the later stage of
600 gestation. However, the current study provides a first view on the genetic effects on
601 metabolites during the pregnancy period and we have observed substantial changes in
602 genetic effect, suggesting the metabolites already changes in the early period of
603 pregnancy.

604

605 This study consistently proves that NIPT sequencing data can be used for medical
606 genetic studies following our previous study introducing the utility of the non-
607 invasive prenatal sequencing data in human genetics⁵³. As NIPT sequencing is
608 quickly expanding to more than ten million individuals around the world nowadays,
609 sophisticated study designs and suitable methods will enable the use of this invaluable
610 genetic resource. The methods and knowledge obtained from this study can speed up
611 future efforts from the medical and scientific community to integrate multi-omics data
612 with NIPT data to answer fundamental biological and clinical questions.

613 **Acknowledgements**

614 We thank all the patients who participated in this study. We would like to particularly
615 thank Professor Anders Albrechtsen from University of Copenhagen and Professor
616 Rasmus Nielsen from University of California, San Francisco for helpful suggestions
617 on the study. We also thank Shujia Huang, Xiao Liu and Fei Wang for their
618 contributions on this study. The study was supported by National Natural Science
619 Foundation of China (31900487, 81830041), Natural Science Foundation of
620 Guangdong Province, China (2017A030306026), Shenzhen Key Laboratory of
621 Genomics (CXB200903110066A) and Guangdong Enterprise Key Laboratory of
622 Human Disease Genomics (2011A060906007), Guangdong Provincial Key

623 Laboratory of Genome Read and Write(2017B030301011) and Guangdong Provincial
624 Academician Workstation of BGI Synthetic Genomics(2017B090904014).

625 **Author contributions**

626 Conceptualization: Siyang Liu, Xin Jin, Siqi Liu, Liang Lin, Fengxiang Wei, Jianmin
627 Niu, and Rong Wang;
628 Methodology: Siyang Liu, Yan Li;
629 Validation: Yan Li, Chao Nie, Fengxia Su, Ya Gao, Fang Chen;
630 Investigation: Siyang Liu, Jilong Yao, Liang Lin, Xianmei Lan, Linlin Wu, Nannan
631 Kong, Yuqing Deng, Yan Li, Jiansheng Xie, Zilong Li, Likuan Xiong, Jinghui Ren,
632 Xuemei Qiu, Weihua Zhao, Ya Gao, Yuanqing Chen, Fengxia Su, Yun Zhou,
633 Weiqiao Rao, Jing Zhang, Guixue Hou, Liping Huang, Xinhong Liu, Chao Nie,
634 Liqiong Luo, Zengyou Liu, Fang Chen, Shengmou Lin, Lijian Zhao, Dan Jiang, Ye
635 Yin, Xun Xu, Jian Wang, Huanming Yang, Rong Wang, Fengxiang Wei, Xin Jin ,
636 Siqi Liu;
637 Formal Analysis: Siyang Liu, Xianmei Lan, Zilong Li, Nannan Kong, Yan Li;
638 Data Curation, Nannan Kong, Yuan Wang, Xiaoxia Wu;
639 Writing-Original Draft: Siyang Liu;
640 Writing-Review & Editing: Siyang Liu, Siqi Liu;
641 Visualization: Siyang Liu, Xianmei Lan, Zilong Li, Nannan Kong;
642 Supervision: Siqi Liu, Jian Wang, Huangming Yang;
643 Project Administration: Liang Lin;
644 Funding Acquisition: Xin Jin, Siyang Liu;

645 **Methods**

646 **Participant recruitment and plasma sample preparation**

647 All the 34, 394 participants were recruited via the non-invasive fetal trisomy test and
648 the pregnancy nutrition program between year 2017 and 2018. They underwent
649 pretest counseling and filled in informed written consent before blood sampling.
650 The study was reviewed and approved by the Institutional Review Board of BGI
651 (BGI-IRB21184) in strict compliance with regulations regarding ethical
652 considerations and personal data protection.

653 5 μ g peripheral whole blood was drawn from each of the participant and stored
654 in the EDTA anticoagulant tubes to avoid hemolysis. The plasma was obtained by
655 centrifugation (3000 rpm, 10 min) and was preserved at -80°C until assay. As for
656 amino acid extraction, 40 μ L plasma was mixed with 20 μ L stable-isotope labeled
657 internal standard (IS) in sulfosalicylic acid to precipitate proteins, followed by vortex
658 and centrifugation (4000 rpm, 4°C, 20 min). Regarding hormones, 250 μ L plasma
659 mixed with 205 μ L IS solutions were filtered through solid-phase extraction (SPE),
660 followed by washing with 25% methanol and eluting with 100% dichloromethane.
661 The elution was evaporated by nitrogen and was reconstituted in 25% methanol. For
662 vitamin extraction, approximately 200 μ L plasma were mixed with 600 μ L
663 methanol/acetic acid/IS for water-soluble vitamins (WSV), or with 1000 μ L
664 methanol/acetonitrile/IS for fat-soluble vitamins (FSV), followed by vortex and
665 centrifugation (4000 g, 4°C, 20 min). The supernatants were evaporated by nitrogen
666 and were reconstituted by either 60 μ L deionized water (WSV) or 80 μ L 80%
667 acetonitrile (FSV). As regards metal elements, 100 μ L plasma were mixed with 400
668 μ L nitric acid/hydrogen peroxide for decomposition at 105°C for 3 hours. The

669 decomposed products were cooled down and delivered to mass spectrometry after
670 dilution with deionized water.

671

672 **Quantitative measurement of metabolites using mass spectrometry**

673 The metabolites in blood were targeted and quantified using liquid chromatography
674 coupled with triple quadrupole mass spectrometry (LC MS/MS), including
675 ACQUITY UPLC I-Class (Waters) mounted with C18 column, Triple Quad 5500
676 (Sciex) and Xevo TQ-S (Waters). The MS/MS spectra corresponding to metabolites
677 were acquired at positive ion mode with multiple reaction monitoring scans. The
678 metal elements were measured by inductively coupled plasma mass spectrometry
679 (ICP MS/MS), i.e. 7700x ICP-MS (Agilent). The mass spectra acquired were
680 processed with MultiQuant (V. 3.0.2, Sciex) for amino acids, hormones and FSV,
681 MassLynx (V. 4.1, Waters) for WSV and MassHunter (V. B.01.03, Agilent) for metal
682 elements. The calibration curves were implemented with stable isotope-labeled
683 compounds as internal standards. The accuracy of quality controls (QCs) with
684 isotope-labeled IS was managed approximately every 15 samples to ensure the inter-
685 batch stability.

686 **Sequencing assays**

687 Details of the sequencing protocol were published previously in ⁵⁴. In brief, within 8h
688 of blood collection, plasma was extracted from whole blood after two turns of
689 centrifugation. The plasma samples were subsequently subjected to library
690 construction, sample quality control and 36-cycle single-end multiplex sequencing on
691 BlackBird or BGI-seq500 platform. The reads were trimmed to 35bp before
692 bioinformatic analysis. Filtering of poor quality reads was carried out using
693 SOAPnuke (<https://github.com/BGI-flexlab/SOAPnuke>). A read was removed if it
694 contained more than 30% low quality bases ($Q \leq 2$) or N bases. In general, each
695 participant was whole-genome sequenced to 5-10 million cleaned reads, representing
696 a sequencing depth around 0.06x -0.1x.

697 **Correlation between metabolite concentrations**

698 We have investigated the correlation between metabolite concentrations in two
699 aspects. First, we estimated the spearman correlation coefficient between the
700 phenotypes. Second, we applied the correlation model implemented in LD score
701 correlation (LDSC) to compute the genetic correlation between any two of the
702 metabolites (Bulik-Sullivan et al., 2015). Correlation was visualized as a heatmap
703 using R.

704 **Genotype imputation**

705 We employed STITCH (version 1.2.7) ⁵⁵ to impute genotype probabilities for the BB
706 and Seq500 individuals in a five-megabase window with a 250K buffer assuming 40
707 ancestral haplotypes, respectively. Allele frequency information from the Chinese
708 population (CHB+CHS+CDX, N=301) in the 1KG impute2 reference panel was used
709 for the initial values for the EM optimization of the model parameters. The imputed
710 variants included 8.16 million known polymorphic sites in 22 autosomal
711 chromosomes and chrX with a 1KG East Asian allele frequency ≥ 0.01 . All the loci
712 recorded in the GWAS catalog are also included for imputation. For each of the
713 imputed sites, there is an IMPUTE2-style info score (Marchini et al., 2007) and a P-
714 value for violation of Hardy Weinberg equilibrium (HWE-pvalue in short)
715 (Wigginton et al., 2005). We used info score greater than 0.4, and minor allele
716 frequency greater than 0.01 as filtration threshold to obtain the significantly

717 associated variants. Imputation accuracies were estimated using Pearson's R2
718 between the true genotype from high-coverage WGS of 50 participants (40x) and the
719 imputed genotype dosage.

720 **Genome-wide association analysis with PLINK**

721 Since the variance regression approach assumes that the trait values follow
722 multivariate normal distribution, we performed a rank transformation of the
723 metabolite raw values using the empirical normal quantile transformation approach
724 after removal of outliers in the raw value distribution⁵⁶. We then applied the linear
725 regression model implemented in Plink v1.9 to detect the associations between the
726 imputed genotype dosage and the transformed values of the metabolites⁵⁷. The
727 covariates in the regression model include maternal age, gestational week upon
728 sampling, the top three principal components of PCA and the inferred gender of the
729 fetus computed from the NIPT data. The regression analysis was performed for black
730 bird and Seq500 sequencing data independently that produced effect size, standard
731 errors, number of effect individuals and p-value for each of the imputed sites.

732 **Meta-analysis and multi-trait genome-wide association test using the MTAG 733 approach**

734 We applied the inverse variance based approach implemented in METAL to perform
735 the meta-analysis integrating the effects from BB and Seq500 regression outcomes²⁶.
736 Since many metabolites may have shared genetic correlation, we applied the multi-
737 trait genome-wide association studies based on the metal summary statistics using the
738 MTAG approach²⁷. The resulting beta, p-value and standard errors were used for
739 downstream analysis. During the MTAG analysis, we noticed that the time and
740 memory that required for performing multi-trait genome wide association studies
741 using all the 84 metabolites is too consuming due to very large Omega and Sigma
742 matrices and we cannot finish the computation successfully in the end. After a
743 personal communication with the MTAG author, we performed the MTAG analysis
744 for each amino acid (N=37), element (N=24), hormone (N=13) and vitamin (N=10),
745 respectively.

746 **Identification of independent significant loci**

747 After learning the effect size and the p-value from the meta-analysis, we defined
748 independent loci as significant variants clustered in a 1Mbp window. The lead SNP
749 was defined as the SNP in the 1Mbp window that has the smallest p-value. We didn't
750 observe multiple independent signals in one locus after performing a conditional test
751 using Plink v1.9. Furthermore, locuszoom was applied to visualize the loci⁵⁸. The
752 reported loci were determined from the conditional test after the single marker
753 analysis using a significance threshold P value $\leq 5 \times 10^{-8}$. Since the genome-wide
754 association study was performed on the quantile-transformed metabolite value (β_{nor}),
755 we applied the following formula $\beta = \beta_{nor} * sd_{pheno}$ to obtain the effect on real
756 metabolite level where sd_{pheno} refers to the standard deviation of values of certain
757 phenotype. Both β_{nor} and β were reported in Table 1 and Table S3.

758 **SNP heritability**

759 The genomic inflation factor λ_{GC} , the heritability, the intercept, and ratio using the
760 LD score regression approach based on the summary statistics³⁵. 12 out of the 84
761 metabolites display negative heritability. After personal communication with the
762
763

764 LDSC author, this may be due to a lack of power for those traits because of small
765 sample size. We set those metabolites with negative heritability as zero in Figure S8
766 and Table S7. The UK Biobank heritability was obtained from
767 https://nealelab.github.io/UKBB_ldsc/ .

768
769

770 **Comparison with the non-pregnancy population**

771 1,553 participants (Female N=642 , Male N=911) from Shenzhen local area were
772 recruited. Written form of consent was signed by each individual. The white blood
773 cell of each individual were whole-genome sequenced to around 30x using the BGI-
774 seq500 platform. A total of the 81 out of the 84 metabolites were investigated for each
775 individual using the same protocol for the pregnancy study that the amino acids,
776 vitamins and hormones were assayed using plasma while the elements were assayed
777 using the whole blood. The three metabolites not investigated were I, Ba and DOC.
778 WGS data were aligned and variants called by the Picard
779 (<http://picard.sourceforge.net>), BWA⁵⁹ and GATK v3.8 best practice⁶⁰ pipeline. SNPs
780 with mapping quality greater than 40, sequencing depth greater than 4, variant quality
781 greater than 2.0, Phred score of Fisher's test p-value for strand bias smaller than 60.0,
782 Haplotype score smaller than 13.0 and distance of alternative allele from the end of
783 reads greater than 8.0 were kept for following analyses. We removed SNPs deviating
784 from Hardy-Weinberg (P-value < 1×10⁻⁵), markers with more than 1% missing
785 genotype data and variants with smaller than 1% minor allele frequencies. Individuals
786 with heterozygosity greater than three standard deviations were excluded. One
787 individual among relatives within 3rd degree of relationship was randomly selected to
788 keep in the clean data set. PCA was performed to investigate population stratification.
789 No clean sub-cluster was observed. Typical north to south Grand aunt was
790 demonstrated by the first principal component. Linear regression adding individual
791 sex and top two principal components as covariates was performed for each of the
792 significant locus identified from the pregnancy study. We also performed regression
793 using only the female individuals adjusting top two PCs. No inflation was observed in
794 this analysis ($\lambda_{GC} \sim 1$).

795

796 **KEGG and pleiotropy analysis**

797 To investigate the pleiotropy among metabolites, for each of the 53 significant loci (P
798 value $\leq 5 \times 10^{-8}$), we visualized the effect and P value for the lead SNP in a forest plot
799 (**Figure S10**). Strong pleiotropy was defined as the observation of a lead SNP that
800 affects at least two metabolites with a significance level of 10⁻³. To further understand
801 the relationship between the SNP variants and multiple metabolites, we conducted a
802 wald ratio estimation implemented in the `tsls` package in R for each of the 27 variants
803 that suggest pleiotropy. We performed a total of 54 tests, corresponding to a p-value
804 less than 0.001 after the Bonferroni correction. We summarized and visualized the
805 pleiotropic correlation in the KEGG pathway using the metaboanalyst website
806 (<https://www.metaboanalyst.ca/>).

807
808

809 **Colocalization analysis**

810 To investigate whether the same genetic variants may drive the associations with
811 metabolites and the pregnancy phenotypes reported in a companion paper, we

812 undertook a colocalization analysis. We applied a stringent Bayesian analysis
813 implemented in the coloc R package⁶¹ with default parameters to estimate the
814 posterior probability (PP) that the metabolites and the pregnancy phenotypes shared a
815 single causal SNP at the locus. The SNPs in the 1 Mb range of the tested
816 instrumental variable that have MAF over 0.05 were used for analysis. Metabolites
817 with a PPH4 > 0.5 (posterior probabilities of two traits sharing one causal SNP) were
818 considered to be colocalized.
819

820 **Two-sample Mendelian randomization between metabolites and 120 traits** 821 **from the Biobank of Japan**

822 To investigate the potential causal impact of the metabolite on human traits or
823 diseases, we performed Mendelian randomization analyses using four methods: the
824 inverse variance weighted (IVW) method, the MR Egger method, the weighted
825 median method, and the weighted mode method implemented in the TwoSampleMR
826 R package⁶². The metabolites that passed the correction threshold $P < 0.01$ were
827 retained for pleiotropy and heterogeneity evaluation. Cochran's Q statistic was used to
828 examine the heterogeneity of the IVs. Horizontal pleiotropy was evaluated using the
829 MR Egger approach with a return of intercept values. MR Steiger test⁶³ was applied to
830 test the directionality of the effect. Only potential causal effects passing the
831 pleiotropy, directionality, and heterogeneity tests were reported in our study ($P > 0.05$).
832

833 **Data availability**

834 The summary statistics of the meta-analysis outcome and the MTAG outcome
835 have been deposited into CNGB Sequence Archive 22 of CNGBdb23 with accession
836 number CNP0003025.
837

838 **Reference**

- 839
- 840 1. Baker, S. A. & Rutter, J. Metabolites as signalling molecules. *Nat. Rev. Mol.*
841 *Cell Biol.* (2023). doi:10.1038/s41580-022-00572-w
 - 842 2. Surendran, P. et al., Rare and common genetic determinants of metabolic
843 individuality and their effects on human health. *Nat. Med.* **28**, 2321–2332
844 (2022).
 - 845 3. Diener, C. et al., Genome–microbiome interplay provides insight into the
846 determinants of the human blood metabolome. *Nat. Metab.* **4**, 1560–1572
847 (2022).
 - 848 4. Veldhoen, M. & Ferreira, C. Influence of nutrient-derived metabolites on
849 lymphocyte immunity. *Nat. Med.* **21**, 709–718 (2015).
 - 850 5. Wishart, D. S. Emerging applications of metabolomics in drug discovery and
851 precision medicine. *Nat. Rev. Drug Discov.* **15**, 473–484 (2016).
 - 852 6. Claussnitzer, M. et al., A brief history of human disease genetics. 179–189
853 (2020).
 - 854 7. MacArthur, J. et al., The new NHGRI-EBI Catalog of published genome-wide
855 association studies (GWAS Catalog). *Nucleic Acids Res.* **45**, D896–D901
856 (2017).

- 857 8. Sirugo, G., Williams, S. M. & Tishkoff, S. A. The Missing Diversity in Human
858 Genetic Studies. *Cell* **177**, 26–31 (2019).
- 859 9. Kim, Y. J. et al., Large-scale genome-wide association studies in east Asians
860 identify new genetic loci influencing metabolic traits. *Nat. Genet.* **43**, 990–997
861 (2011).
- 862 10. Smith, P. H., Bessette, A. J., Weinberger, A. H., Sheffer, C. E. & Mckee, S. A.
863 Novel Association of Genetic Markers Affecting CYP2A6 activity and Lung
864 Cancer Risk. *Physiol. Behav.* **92**, 135–140 (2016).
- 865 11. Duff Putu, Jean Shoveller., Julio Montaner., Cindy Feng., Rachel Nicoletti.,
866 Kate Shannon ., G. O. et al., Genetic Determinants of 1,3-Butadiene
867 Metabolism and Detoxification in Three Populations of Smokers with Different
868 Risks of Lung Cancer. *Physiol. Behav.* **176**, 139–148 (2016).
- 869 12. Chai, J. F. et al., Associations with metabolites in Chinese suggest new
870 metabolic roles in Alzheimer’s and Parkinson’s diseases. *Hum. Mol. Genet.* **29**,
871 189–201 (2020).
- 872 13. Wang, Z. et al., Genome-wide association study of metabolites in patients
873 with coronary artery disease identified novel metabolite quantitative trait loci.
874 *Clin. Transl. Med.* **11**, 1–6 (2021).
- 875 14. Li, S. et al., Identification of common genetic variants associated with serum
876 concentrations of p, p'-DDE in non-occupational populations in eastern China.
877 *Environ. Int.* **152**, 106507 (2021).
- 878 15. Kanai, M. et al., Genetic analysis of quantitative traits in the Japanese
879 population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–
880 400 (2018).
- 881 16. Wu, Q. et al., Multi-stage metabolomics and genetic analyses identified
882 metabolite biomarkers of metabolic syndrome and their genetic determinants.
883 *EBioMedicine* **74**, (2021).
- 884 17. Liang, L. et al., Metabolic Dynamics and Prediction of Gestational Age and
885 Time to Delivery in Pregnant Women. *Cell* **181**, 1680-1692.e15 (2020).
- 886 18. Duerbeck, N. B. & Dowling, D. D. Vitamin A: Too much of a good thing?
887 *Obstetrical and Gynecological Survey* (2012).
888 doi:10.1097/OGX.0b013e318244c52d
- 889 19. Prevention of neural tube defects: Results of the Medical Research Council
890 Vitamin Study. *Lancet* (1991). doi:10.1016/0140-6736(91)90133-A
- 891 20. Mistry, H. D., Broughton Pipkin, F., Redman, C. W. G. & Poston, L. Selenium
892 in reproductive health. *American Journal of Obstetrics and Gynecology* (2012).
893 doi:10.1016/j.ajog.2011.07.034
- 894 21. Rayman, M. P., Bode, P. & Redman, C. W. G. Low selenium status is
895 associated with the occurrence of the pregnancy disease preeclampsia in
896 women from the United Kingdom. *Am. J. Obstet. Gynecol.* (2003).
897 doi:10.1067/S0002-9378(03)00723-3
- 898 22. Zeghoud, F. et al., Subclinical vitamin D deficiency in neonates: Definition
899 and response to vitamin D supplements. *Am. J. Clin. Nutr.* (1997).
900 doi:10.1093/ajcn/65.3.771
- 901 23. Liu, S. et al., Genomic Analyses from Non-invasive Prenatal Testing Reveal
902 Genetic Associations , Patterns of Viral Infections , and Chinese Population
903 History Article Genomic Analyses from Non-invasive Prenatal Testing Reveal
904 Genetic Associations , Patterns of Viral Inf. *Cell* **175**, 347-359.e14 (2018).
- 905 24. Jiang, F. et al., Noninvasive Fetal Trisomy (NIFTY) test: An advanced
906 noninvasive prenatal diagnosis methodology for fetal autosomal and sex

- 907 chromosomal aneuploidies. *BMC Med. Genomics* **5**, (2012).
- 908 25. Chang, C. C. et al., Second-generation PLINK: Rising to the challenge of
909 larger and richer datasets. *Gigascience* **4**, (2015).
- 910 26. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-
911 analysis of genomewide association scans. *Bioinformatics* (2010).
912 doi:10.1093/bioinformatics/btq340
- 913 27. Turley, P. et al., Multi-trait analysis of genome-wide association summary
914 statistics using MTAG. *Nat. Genet.* (2018). doi:10.1038/s41588-017-0009-4
- 915 28. Stelzer, G. et al., The GeneCards suite: from gene data mining to disease
916 genome sequence analyses. *Curr. Protoc. Bioinforma.* **54**, 1–30 (2016).
- 917 29. Kamat, M. A. et al., PhenoScanner V2: An expanded tool for searching human
918 genotype-phenotype associations. *Bioinformatics* **35**, 4851–4853 (2019).
- 919 30. Buniello, A. et al., The NHGRI-EBI GWAS Catalog of published genome-
920 wide association studies, targeted arrays and summary statistics 2019. *Nucleic
921 Acids Res.* **47**, D1005–D1012 (2019).
- 922 31. Melville, G. W., Siegler, J. C. & Marshall, P. W. M. The effects of d-aspartic
923 acid supplementation in resistance-trained men over a three month training
924 period: A randomised controlled trial. *PLoS One* **12**, 1–20 (2017).
- 925 32. Smith, A. D. & Refsum, H. Homocysteine – from disease biomarker to disease
926 prevention. *J. Intern. Med.* **290**, 826–854 (2021).
- 927 33. Seppälä, I. et al., Genome-wide association study on dimethylarginines reveals
928 novel AGXT2 variants associated with heart rate variability but not with
929 overall mortality. *Eur. Heart J.* (2014). doi:10.1093/eurheartj/eh447
- 930 34. Stelzer, G. et al., The GeneCards suite: From gene data mining to disease
931 genome sequence analyses. *Curr. Protoc. Bioinforma.* (2016).
932 doi:10.1002/cpbi.5
- 933 35. Bulik-Sullivan, B. K. et al., LD Score regression distinguishes confounding
934 from polygenicity in genome-wide association studies. *Nat. Genet.* (2015).
935 doi:10.1038/ng.3211
- 936 36. Kettunen, J. et al., Genome-wide study for circulating metabolites identifies 62
937 loci and reveals novel systemic effects of LPA. *Nat. Commun.* **7**, 1–9 (2016).
- 938 37. Nie, C. et al., Distinct biological ages of organs and systems identified from a
939 multi-omics study. *Cell Rep.* **38**, 110459 (2022).
- 940 38. Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K. & Tanabe, M. New
941 approach for understanding genome variations in KEGG. *Nucleic Acids Res.*
942 (2019). doi:10.1093/nar/gky962
- 943 39. Burgess, S., Small, D. S. & Thompson, S. G. A review of instrumental variable
944 estimators for Mendelian randomization. *Stat. Methods Med. Res.* **26**, 2333–
945 2355 (2017).
- 946 40. Lyon, M. S. et al., The variant call format provides efficient and robust storage
947 of GWAS summary statistics. *Genome Biol.* **22**, 1–10 (2021).
- 948 41. Husemoen, L. L. N. et al., Mendelian randomisation study of the associations
949 of vitamin B12 and folate genetic risk scores with blood pressure and fasting
950 serum lipid levels in three Danish population-based studies. *Eur. J. Clin. Nutr.*
951 **70**, 613–619 (2016).
- 952 42. Holmes, M. V. et al., Effect modification by population dietary folate on the
953 association between MTHFR genotype, homocysteine, and stroke risk: A meta-
954 analysis of genetic studies and randomised trials. *Lancet* **378**, 584–594 (2011).
- 955 43. Balashova, O. A., Visina, O. & Borodinsky, L. N. Folate action in nervous
956 system development and disease. *Dev. Neurobiol.* **78**, 391–402 (2018).

- 957 44. Xu, M. Y. et al., Vitamin D and graves' disease: A meta-analysis update.
958 *Nutrients* **7**, 3813–3827 (2015).
- 959 45. Grove-Laugesen, D. et al., Effect of 9 months of vitamin D supplementation
960 on arterial stiffness and blood pressure in Graves' disease: a randomized
961 clinical trial. *Endocrine* **66**, 386–397 (2019).
- 962 46. Cho, Y. Y. & Chung, Y. J. Vitamin D supplementation does not prevent the
963 recurrence of Graves' disease. *Sci. Rep.* **10**, 1–7 (2020).
- 964 47. Ye, Y., Yang, H., Wang, Y. & Zhao, H. A comprehensive genetic and
965 epidemiological association analysis of vitamin D with common diseases/traits
966 in the UK Biobank. *Genet. Epidemiol.* **45**, 24–35 (2021).
- 967 48. Raffetti, E. et al., Cortisol Concentration as Predictor of Tobacco Initiation in
968 Adolescents: Results From a Population-Based Swedish Cohort. *J. Adolesc.*
969 *Heal.* **68**, 758–764 (2021).
- 970 49. Masi, A. T. et al., Lower serum androstenedione levels in pre-rheumatoid
971 arthritis versus normal control women: Correlations with lower serum cortisol
972 levels. *Autoimmune Dis.* **2013**, (2013).
- 973 50. Buniello, A. et al., The NHGRI-EBI GWAS Catalog of published genome-
974 wide association studies, targeted arrays and summary statistics 2019. *Nucleic*
975 *Acids Res.* (2019). doi:10.1093/nar/gky1120
- 976 51. Qiu, X. et al., The Born in Guangzhou Cohort Study (BIGCS). *Eur. J.*
977 *Epidemiol.* **32**, 337–346 (2017).
- 978 52. Smith, G. D. & Hemani, G. Mendelian randomization: Genetic anchors for
979 causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**, 89–98
980 (2014).
- 981 53. Liu, S. et al., Genomic Analyses from Non-invasive Prenatal Testing Reveal
982 Genetic Associations, Patterns of Viral Infections, and Chinese Population
983 History. *Cell* (2018). doi:10.1016/j.cell.2018.08.016
- 984 54. Zhang, H. H. et al., Noninvasive Prenatal Testing for Trisomy 21, 18 and 13 -
985 Clinical Experience from 146,958 Pregnancies. *Ultrasound Obstet. Gynecol.*
986 n/a-n/a (2015). doi:10.1002/uog.14792
- 987 55. Davies, R. W., Flint, J., Myers, S. & Mott, R. Rapid genotype imputation from
988 sequence without reference panels. *Nat. Genet.* **48**, 1–7 (2016).
- 989 56. Ren, S. et al., RNA-seq analysis of prostate cancer in the Chinese population
990 identifies recurrent gene fusions, cancer-associated long noncoding RNAs and
991 aberrant alternative splicings. *Cell Res.* **22**, 806–21 (2012).
- 992 57. Chang, C. C. et al., Second-generation PLINK: rising to the challenge of larger
993 and richer datasets. *Gigascience* **4**, 7 (2015).
- 994 58. Pruim, R. J. et al., LocusZoom: Regional visualization of genome-wide
995 association scan results. *收入 Bioinformatics* **27**, 2336–2337 (2011).
- 996 59. Li, H. Aligning sequence reads, clone sequences and assembly contigs with
997 BWA-MEM. *arXiv Prepr. arXiv1303.3997* (2013).
- 998 60. Depristo, M. A. et al., A framework for variation discovery and genotyping
999 using next-generation DNA sequencing data. *Nat. Genet.* (2011).
1000 doi:10.1038/ng.806
- 1001 61. Giambartolomei, C. et al., Bayesian Test for Colocalisation between Pairs of
1002 Genetic Association Studies Using Summary Statistics. *PLoS Genet.* **10**,
1003 (2014).
- 1004 62. Hemani, G. et al., The MR-base platform supports systematic causal inference
1005 across the human phenome. *Elife* **7**, 1–29 (2018).
- 1006 63. Hemani, G., Tilling, K. & Smith, G. D. Orienting the causal relationship

1007 between imprecisely measured traits using genetic instruments. *PLOS Genet.*
1008 **13**, e1007081 (2017).
1009
1010

Figures and tables

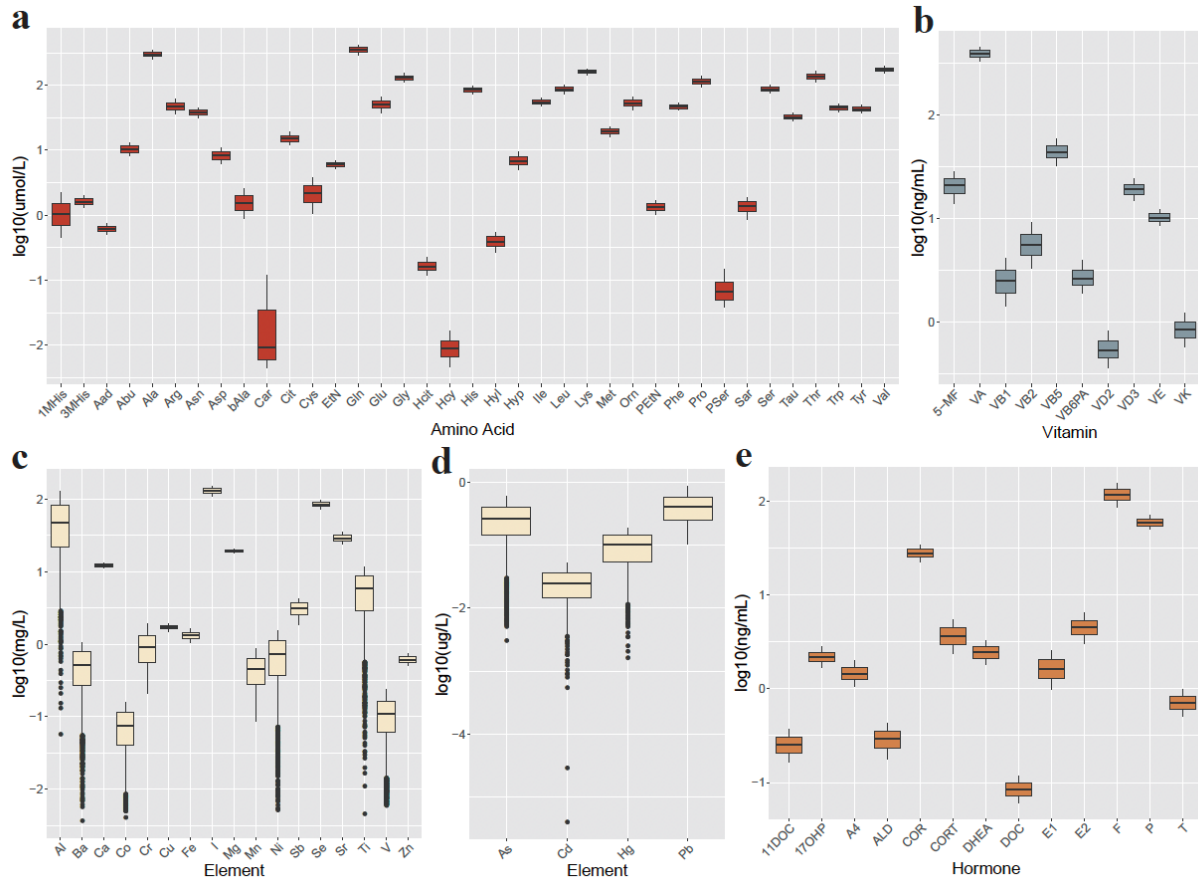


Figure 1. Phenotypic distribution of the 84 metabolites

Shown were the logarithm of the concentration for a) 37 amino acids b) 10 vitamins c) and d) 24 elements and e) 13 hormones. Units of the metabolites were shown in the title of the y-axis. The outliers were excluded in the genome-wide association analysis.

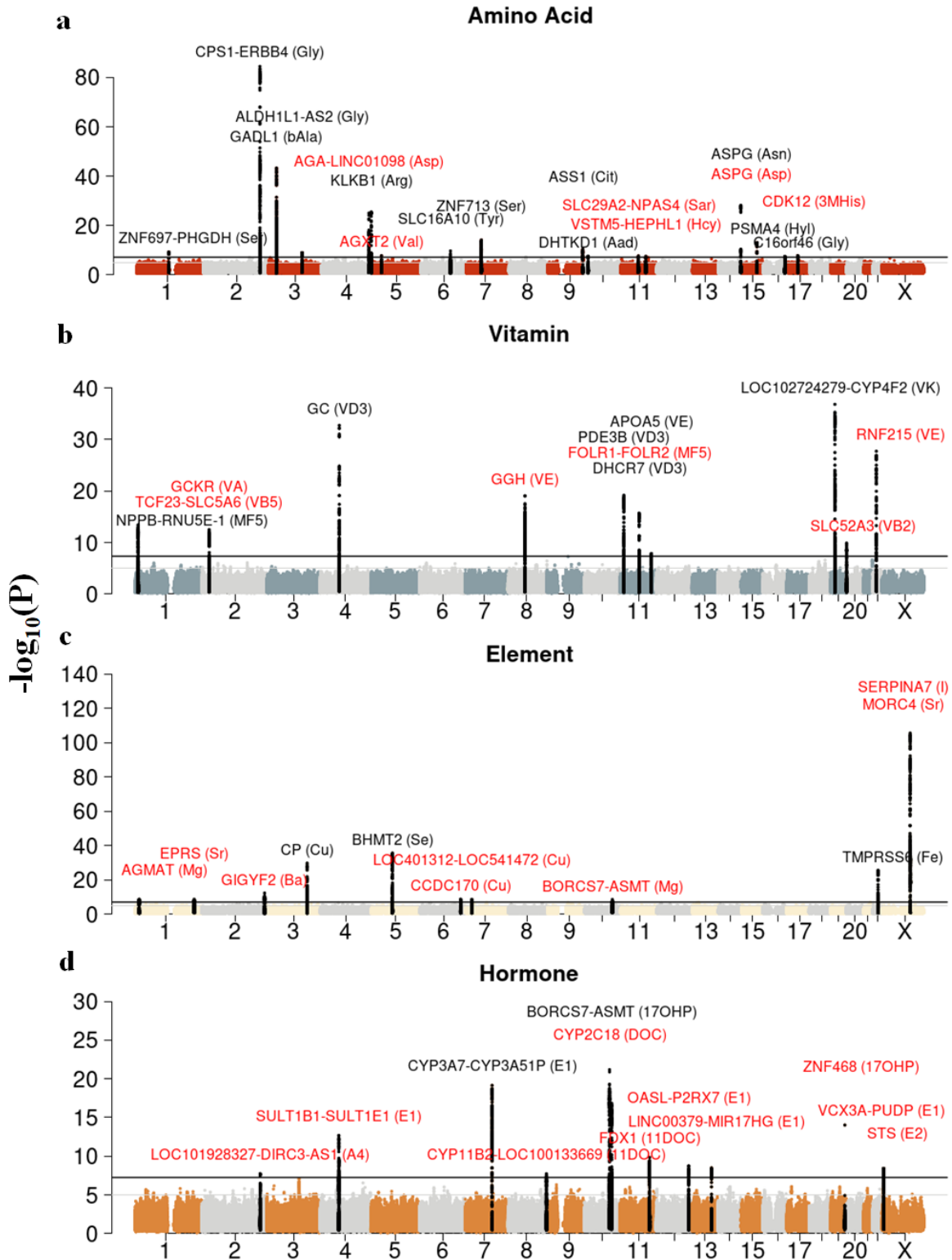


Figure 2. The 53 significant genetic associations with the metabolites. Manhattan plot of all the genetic associations for amino acids (a), vitamins (b), metal elements (c) and hormones

(d). The x-axis is the chromosomal position and the y-axis represents the minus logarithmic transformation of the P value from the GWAS regression model. The black and grey horizontal line represents the P value of the genome-wide significance level at $P=5.0 \times 10^{-8}$ and $P=1.0 \times 10^{-5}$. Each significant locus is annotated by the nearest gene symbol with the corresponding trait in parenthesis. Gene symbols in red suggest that the genetic association locus (1Mbp centering on the lead SNP) was not previously identified in the GWAS catalog or PubMed. Gene symbols in black reflect known associations.

It is made available under a [CC-BY-ND 4.0 International license](https://creativecommons.org/licenses/by-nd/4.0/).

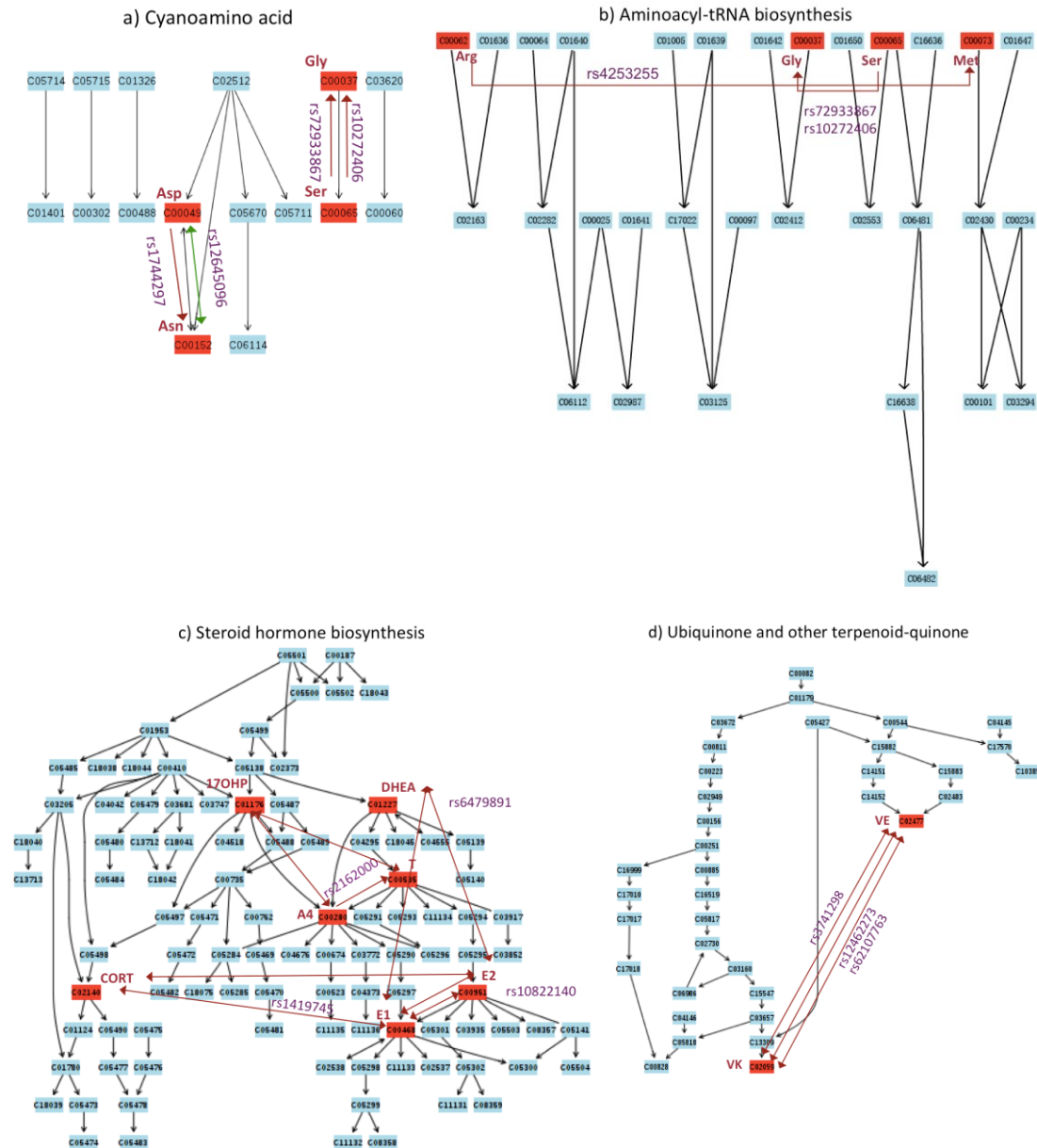


Figure 3. Genetic variants that affect multiple metabolites in the context of KEGG pathway

Shown are SNPs that suggest a significant causal effect in the wald ratio estimates ($p < 0.05/53$). Red lines suggest the same direction of the effect allele while the green line indicates the opposite direction of the effect alleles for the two connected metabolites. Two-edged arrows suggest significant effects of the SNP were observed on both connected metabolites while the single-edged arrows suggest a significant effect on the pointed metabolite. Statistic details of the wald ratio estimates can be found in Table S8. KEGG pathway plot was generated using the online pathway analysis module at <https://www.metaboanalyst.ca/>.

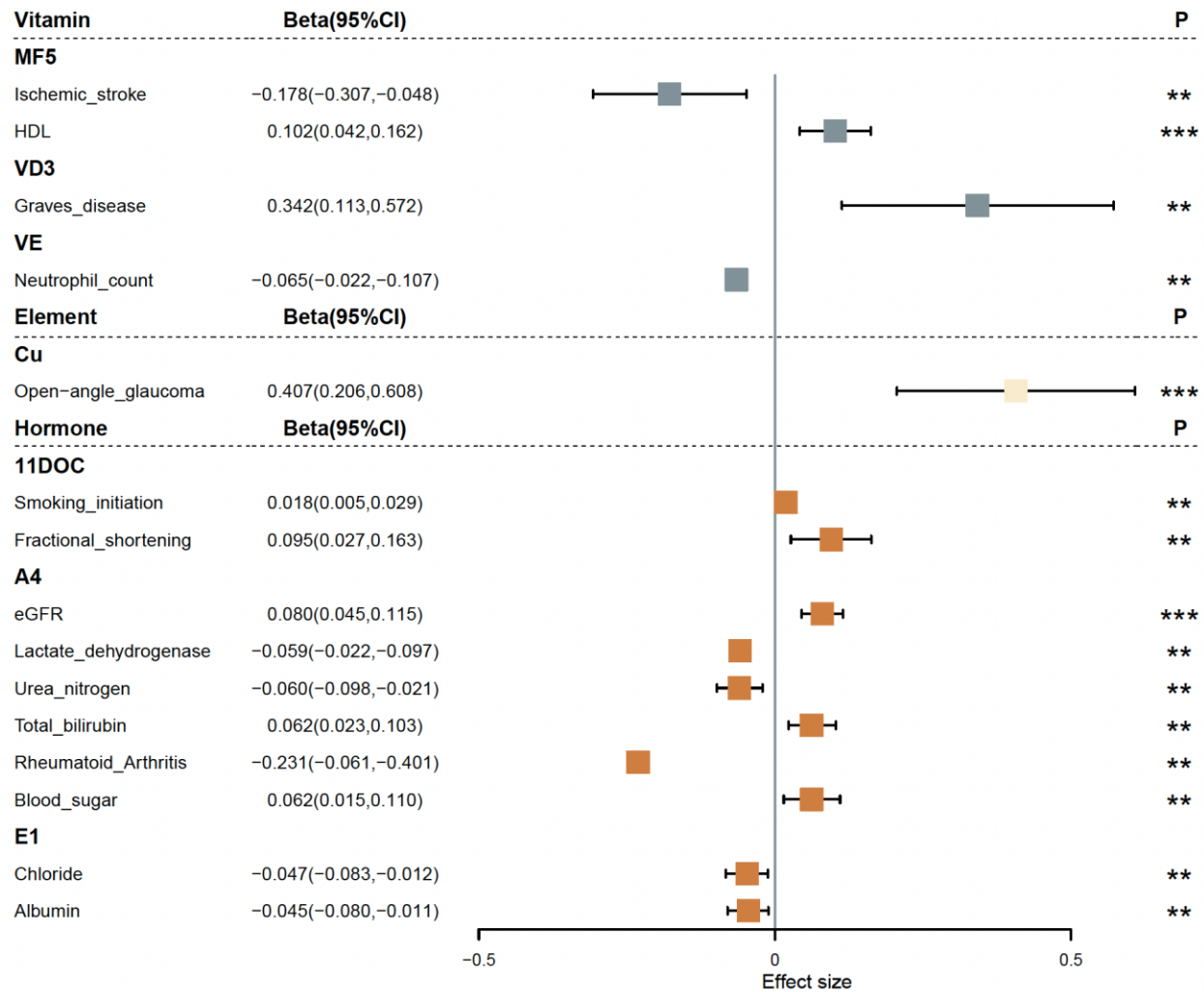


Figure 4. Forest plot displaying the effect of the pregnant metabolites on adult traits by two-sample mendelian randomization. Effects and P values were shown if $P < 0.01$ based on a Bonferroni correction of five categories of metabolites. The colors of the boxes correspond to the metabolite categories in Figures 1 and 2. The inverse variance weighted method was applied for the MR analysis and only potential causal relationships without horizontal pleiotropic effect ($P > 0.05$) and heterogeneity ($P > 0.05$) were shown. **: $P < 0.01$; ***: $P < 0.001$

Table 1. The fifty-three genetic loci associated with maternal plasma metabolites

| Metabolites | Effect allele | Gene | $\beta_{\text{nor}}^{\#1}$ | $\beta^{\#2}$ | <i>P</i> | Novelty |
|---|---------------|------------------------------------|----------------------------|---------------|-----------|---------|
| Amino acids ($\mu\text{mol/L}$) | | | | | | |
| 3MHis | rs11078912-A | <i>CDK12</i> | 0.11 | 0.18 | 1.88E-08 | Novel |
| Aad | rs2062987-A | <i>DHTKD1</i> | -0.11 | -0.07 | 3.01E-08 | Known |
| Arg | rs4253311-A | <i>KLKB1</i> | -0.20 | -9.99 | 3.76E-26 | Known |
| Asn | rs11627783-T | <i>ASPG</i> | 0.30 | 11.37 | 6.61E-29 | Known |
| Asp | rs12645096-T | <i>AGA-LINC01098</i> | -0.21 | -1.80 | 9.07E-26 | Novel |
| Asp | rs1744294-T | <i>ASPG</i> | -0.17 | -1.48 | 1.55E-16 | Novel |
| bAla | rs6550015-T | <i>GADL1</i> | 0.21 | 0.40 | 4.98E-44 | Known |
| Cit | rs4740223-A | <i>ASS1</i> | -0.16 | -2.56 | 4.27E-11 | Known |
| Gly | rs72933867-T | <i>CPS1-ERBB4</i> | -0.46 | -60.92 | 3.26E-85 | Known |
| Gly | rs35331719-A | <i>ALDH1L1-AS2</i> | 0.10 | 13.02 | 1.10E-09 | Known |
| Gly | rs804896-A | <i>C16orf46</i> | -0.09 | -11.92 | 3.18E-08 | Known |
| Hyl | rs1979905-A | <i>PSMA4</i> | 0.16 | 0.07 | 8.81E-14 | Known |
| Hcy | rs16919711-A | <i>VSTM5-HEPHL1</i> | -0.11 | 0.00 | 2.90E-08 | Novel |
| Sar | rs11227478-A | <i>SLC29A2-NPAS4</i> | 0.11 | 0.15 | 2.08E-08 | Novel |
| Ser | rs511059-T | <i>ZNF697-PHGDH</i> | 0.27 | 23.45 | 6.50E-10 | Known |
| Ser | rs10272406-A | <i>ZNF713</i> | 0.15 | 13.50 | 7.04E-15 | Known |
| Tyr | rs14399-A | <i>SLC16A10</i> | -0.10 | -4.15 | 2.06E-10 | Known |
| Val | rs191495-T | <i>AGXT2</i> | -0.07 | -11.66 | 1.69E-08 | Novel |
| Elements ($\mu\text{g/L}$) | | | | | | |
| Ba | rs13424351-A | <i>GIGYF2</i> | -0.10 | -0.09 | 4.32E-13 | Novel |
| I | rs1804495-A | <i>SERPINA7</i> | -0.34 | -43.75 | 2.13E-106 | Novel |
| Se | rs2909855-A | <i>BHMT2</i> | 0.19 | 16.10 | 2.00E-36 | Known |
| Sr | rs2647437-T | <i>EPRS</i> | 0.10 | 2.88 | 1.71E-09 | Novel |
| Sr | rs6616617-T | <i>MORC4</i> | 0.19 | 5.76 | 3.40E-39 | Novel |
| Elements (mg/L) | | | | | | |
| Cu | rs17838831-T | <i>CP</i> | -0.18 | -0.31 | 1.46E-30 | Known |
| Cu | rs9397433-A | <i>CCDC170</i> | 0.10 | 0.18 | 1.60E-09 | Novel |
| Cu | rs58879058-T | <i>LOC401312- LOC541472</i> | 0.10 | 0.16 | 1.97E-09 | Novel |
| Fe | rs877908-T | <i>TMPRSS6</i> | -0.19 | -0.26 | 2.41E-26 | Known |
| Mg | rs6679229-A | <i>AGMAT</i> | -0.08 | -1.51 | 2.00E-09 | Novel |
| Mg | rs12764049-A | <i>BORCS7-ASMT</i> | -0.06 | -1.24 | 2.79E-09 | Novel |
| Hormones (ng/mL) | | | | | | |
| A4 | rs2162000-A | <i>LOC101928327- DIRC3-AS1</i> | -0.078 | -0.112 | 1.79E-08 | Known |

| | | | | | | |
|-------------------------|---------------|----------------------------------|-------|-------|----------|-------|
| DOC | rs1926711-A | <i>CYP2C18</i> | -0.15 | -0.01 | 6.74E-22 | Novel |
| E1 | rs1220712-T | <i>SULT1B1-SULT1E1</i> | 0.10 | 0.18 | 2.11E-13 | Novel |
| E1 | rs4646462-A | <i>CYP3A7-CYP3A51P</i> | 0.18 | 0.34 | 7.20E-20 | Known |
| E1 | rs2708101-T | <i>OASL-P2RX7</i> | 0.11 | 0.21 | 1.84E-09 | Novel |
| E1 | rs74241688-A | <i>LINC00379-MIR17HG</i> | 0.18 | 0.34 | 3.27E-09 | Novel |
| E1 | rs6639741-A | <i>VCX3A-PUDP</i> | 0.06 | 0.12 | 3.81E-09 | Novel |
| E2 | rs802900-A | <i>STS</i> | 0.02 | 0.10 | 3.78E-08 | Novel |
| 11DOC | rs143728108-T | <i>CYP11B2- LOC100133669</i> | 0.08 | 0.02 | 1.89E-08 | Novel |
| 11DOC | rs1073527-T | <i>FDX1</i> | 0.10 | 0.03 | 1.47E-10 | Novel |
| 17OHP | rs10883790-A | <i>BORCS7-ASMT</i> | 0.12 | 0.27 | 1.45E-17 | Known |
| 17OHP | rs4801940-T | <i>ZNF468</i> | -0.20 | -0.45 | 9.49E-15 | Novel |
| Vitamins (ng/mL) | | | | | | |
| MF5 | rs75014290-T | <i>NPPB-RNU5E-1</i> | -0.13 | -2.78 | 2.31E-14 | Known |
| MF5 | rs35954619-A | <i>FOLR1-FOLR2</i> | -0.15 | -3.19 | 3.00E-09 | Novel |
| VB2 | rs6140149-A | <i>SLC52A3</i> | -0.11 | -0.78 | 1.32E-10 | Novel |
| VB5 | rs1659696-A | <i>TCF23-SLC5A6</i> | 0.10 | 5.02 | 3.93E-13 | Novel |
| VA | rs1260333-A | <i>GCKR</i> | 0.11 | 41.53 | 5.31E-09 | Novel |
| VD3 | rs1352846-A | <i>GC</i> | 0.23 | 4.44 | 6.61E-33 | Known |
| VD3 | rs145497491-A | <i>PDE3B</i> | 0.16 | 3.05 | 1.95E-20 | Known |
| VD3 | rs11546509-A | <i>DHCR7</i> | -0.15 | -2.84 | 2.67E-15 | Known |
| VD3 | rs144131853-A | <i>LINC00441</i> | -0.32 | -6.31 | 1.60E-09 | Novel |
| VK | rs62107763-T | <i>LOC102724279- CYP4F2</i> | -0.27 | -0.25 | 1.18E-37 | Known |
| Vitamins (µg/mL) | | | | | | |
| VE | rs3741298-T | <i>APOA5</i> | -0.11 | -1.12 | 1.39E-08 | Known |
| VE | rs61583714-T | <i>RNF215</i> | 0.18 | 1.91 | 2.31E-28 | Novel |
| VE | rs117494536-A | <i>GGH</i> | 0.15 | 1.61 | 8.16E-20 | Novel |

#1: β_{nor} refers to the effect size for the standardized, quantile-transformed metabolite level, corresponding to the unit change of the metabolite per allele change of the variant.

#2: β refers to the effect size for the unstandardized metabolite level, corresponding to the change of the metabolite per allele change of the variant.